

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

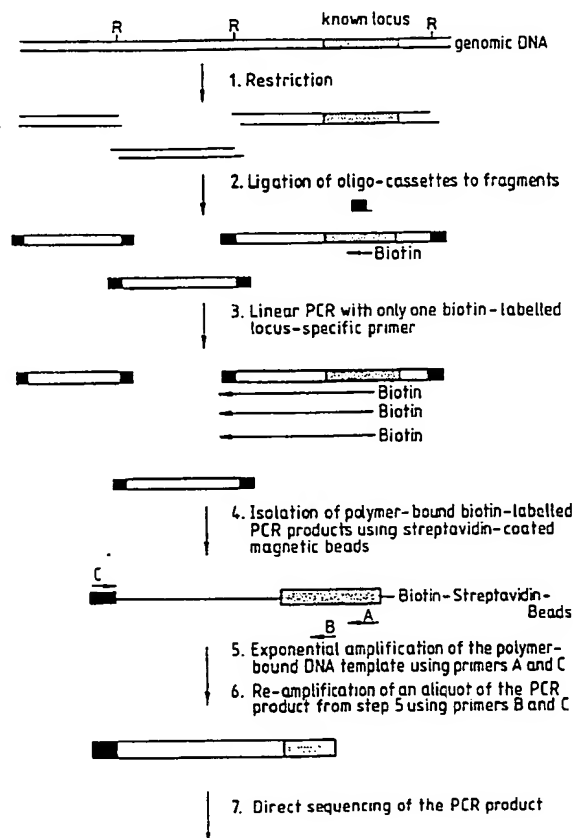
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C12Q 1/68, C07H 21/00</b> <b>// C12P 19/34, C12N 15/10</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 91/18114</b> <b>(43) International Publication Date:</b> 28 November 1991 (28.11.91)
<b>(21) International Application Number:</b> PCT/GB91/00803 <b>(22) International Filing Date:</b> 22 May 1991 (22.05.91) <b>(30) Priority data:</b> 9011454.7 22 May 1990 (22.05.90) GB <b>(71) Applicant (for all designated States except US):</b> MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1 (GB). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only) :</b> JONES, David, Stephen, Charnock [GB/GB]; 96 Cambridge Road, Great Shelford, Cambridge CB2 5JS (GB). ROSENTHAL, Andre [GB/GB]; 8 Hulatt Road, Cambridge CB1 40H (GB). <b>(74) Agents:</b> MERCER, Christopher, Paul et al.; Carpmals & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).		<b>(81) Designated States:</b> AT (European patent), BE (European patent), CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.  <b>Published</b> <i>With international search report.</i>

**(54) Title:** POLYNUCLEOTIDE AMPLIFICATION

**(57) Abstract**

The polynucleotide amplification method described includes the use of a labelled primer that is complementary to a specific known sequence in a target strand. A linear polymerase chain reaction (PCR) step is first conducted with the labelled primer. The labelled linear extension products are then isolated by means of a suitable support matrix that cooperatively binds to the label. The labelled extension products can then be subjected to exponential PCR in the absence of any other strands.



***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TC	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

## POLYNUCLEOTIDE AMPLIFICATION

Field of the Invention

5 The present invention relates to polynucleotide amplification.

Background to the Invention

10 There has been much interest recently in determining the sequence of the human genome. The sequence of many genes and their location within the human genome are already known. It has been proposed that the sequences of unknown areas of the genome could be determined by first determining  
15 the sequence of areas flanking the known genes. In order to do so, it will be necessary to determine the sequence on either side of a known sequence and then, by a series of similar steps, "walk" up or down the genome. In the present specification, a known sequence within a genome is referred  
20 to as a target sequence. Also a fragment of nucleic acid, for instance derived from the human genome, containing such a target sequence is referred to as a target fragment. The method of the present invention allows one to walk up or down a genome starting from a target sequence.

25 In particular, the invention relates to polynucleotide amplification by a cassette mediated polymerase chain reaction technique and to a kit for the same. The term "cassette" means a short section of double stranded (ds)  
30 nucleic acid having a sticky end and a blunt end. The cassette has a known sequence.

The polymerase chain reaction (PCR) is an extremely powerful biochemical technique because it leads to the *in vitro*  
35 production of many copies of a target sequence, thereby avoiding cloning. It is particularly useful for producing acceptable quantities of a target sequence when only a very small amount of the target sequence is naturally available.

**SUBSTITUTE SHEET**

The PCR technique has been used to detect nucleic acid sequences associated with infectious diseases, genetic disorders or cellular disorders such as cancer. A typical  
5 example is the use of the technique in the prenatal diagnosis of sickle cell anaemia using DNA obtained from foetal cells.

The key initial work on the PCR technique was conducted by  
10 Saiki and his co-workers between 1985 and 1987 (Saiki, R. et al. [1985] Science 230 1350 and [1987] Science 239 487). However, despite its recent introduction to biotechnology, it is already a well established technique. A number of patent applications have now been published on the PCR  
15 technique and these include EP-A-0258017, EP-A-0201184, EP-A-0200362 and GB-A-2221909.

In the standard PCR technique, a sample of double stranded nucleic acid (e.g. duplex DNA) having within it a target  
20 sequence is first denatured, usually by heating, so that the two strands of the ds nucleic acid become separated from each other. Two primers are then added. These primers are single stranded oligonucleotides, one of which has a sequence complementary to a region at the 5' end of the  
25 sense strand of the target sequence. The other primer is complementary to a region at the 5' end of the antisense strand of the target sequence. There need not be exact correspondence between the primers and the strand regions. The conditions are then altered to allow the primers to  
30 anneal to their respective strands.

Next, a DNA polymerase (e.g. the Klenow fragment of T4 DNA polymerase, the thermally stable polymerase from *Thermus Aquaticus*, the thermally stable polymerase from *Bacillus*  
35 *Stearothermophilus*, T7 DNA polymerase or modified versions thereof) and the four deoxynucleotide triphosphates are added. Each primer then becomes extended by the synthesis of a new nucleic acid strand complementary to the strand to

which the primer is annealed. Primer extension products are thus formed.

The various strands are then separated from each other,  
5 again by denaturation, and, if necessary, more of the primers are added. Usually, in practice, there is already an excess of primers at the start of the reaction and so further amounts of primers need not be added.

10 The same sequence of reaction steps is then repeated. In repeating the steps, the primers also bind to the previously formed primer extension products resulting in the formation of further nucleic acid strands.

15 By repeatedly recycling the products through the reaction steps, the amount of primer extension products (i.e. the target sequence) increases exponentially.

The PCR technique therefore includes six key steps:

20

1. preparing a solution of single stranded nucleic acid, for example from the denaturation of a double stranded (ds) nucleic acid,
- 25 2. binding a primer at the 5' end of each of the strands of the target sequence,
3. forming double stranded nucleic acid by adding nucleotides to the primers bound on the target  
30 nucleic acid strands by use of a polymerase enzyme,
4. denaturing the double stranded nucleic acid thus formed,
- 35 5. repeating steps 2 to 4, leading to exponential amplification in the production of double stranded nucleic acid, and

6. isolating the prepared nucleic acid.

It will be seen from this that in order to carry out a PCR reaction, the sequence of the target nucleic acid, at least  
5 at its 5' and 3' ends, must be known.

There are many factors which are known to influence the specificity of a given PCR reaction. Some factors include cycling time and temperature, cycling profile of  
10 temperature, PCR buffer quality and strength, additions (e.g. cations) to the PCR buffer, nucleotide triphosphate quality and concentration, DNA polymerase quality and concentration, primer length and concentration. These parameters can all be optimized.

15

Other factors also play an important role in the specificity of a PCR amplification, for example the primary structure of the PCR primers (GC content), formation of secondary  
20 structures during the PCR amplification and complexity of the DNA mixture and concentration of a given template. However, only some of the aforementioned factors have been investigated so far and, in general, little is known about the detailed mechanism of PCR amplification.

25

It is to be noted that in the PCR technique extension of the oligonucleotide primers occurs in a convergent manner relative to the target sequence, i.e. extension in the 5'-  
3' direction occurs within the target sequence.

30

The extension direction in the PCR technique creates a drawback. It only allows amplification of the sections of nucleic acid located between the two primer annealing regions. It does not allow amplification of nucleic acid  
35 sequences that flank the target sequence. Moreover, if the sequence at only one end of the flanking region (i.e. the end adjacent the target sequence) is known, then suitable primers cannot be constructed to enable extension of the

primers towards the target sequence.

Some techniques have been recently developed to overcome this drawback. The first is called the inverse polymerase chain reaction (Triglia, T. et al [1988] NAR 16 8186; Ochman, H. et al [1988] Genetics 120 621). This technique allows the amplification of sections of nucleic acids, even of unknown sequences, that flank a target sequence.

Essentially, the inverse polymerase chain reaction (IPCR) involves a first step of restricting a sample of ds nucleic acid with a restriction endonuclease which forms sticky ends and which does not cut within the target sequence. This restriction produces a number of sticky ended fragments, including one which contains the target sequence and has a flanking sequence of unknown sequence at each side.

The sticky ends of each fragment thus produced are then ligated to each other, resulting in circularisation of the cleaved fragments. In this circularisation step, the fragment containing the target sequence forms a circle wherein the unknown flanking regions form a continuous unknown region connecting the ends of the target sequence. The continuous unknown region can then be subjected to exponential PCR amplification by *inter alia* the addition of the two primers which anneal to the respective ends of the target sequence in such a way that extension takes place into the unknown region. No amplification of circles not containing the target sequence will occur as there will be no site to which the primers can anneal.

The IPCR technique therefore includes the following three key steps:

1. preparing a solution of linear double stranded nucleic acid fragments having sticky ends, one of which fragments contains the target sequence.



2. circularising the nucleic acid fragments, and
3. amplifying the circularised target fragment by the PCR technique.

5

However, there are a number of problems inherent in and encountered with the IPCR technique, largely because circularization occurs only under very specific conditions which have not been investigated in detail and are therefore almost unknown.

Firstly, it is known that the circularisation step (step 2) is dependent on at least two factors, principally the concentration and the size of the nucleic acid fragments.

15

In this regard, it is known that only very dilute solutions of nucleic acid fragments bearing sticky ends can be ligated to form circles. Moreover, the formation of circles is by no means certain and the yield may vary considerably during the ligation step. In particular, the circularization is always accompanied by the formation of linear concatamers of nucleic acid fragments with sticky ends.

The formation of concatamers is due to the fact that each of the excised nucleic acid fragments, whether or not they contain the target sequence, has sticky ends that are complementary not only to each other but also to the ends of all the other fragments produced from the initial nucleic acid digest. The sticky ends of the target fragment can thus ligate not only to themselves but also to the sticky ends of other target fragments or to the sticky ends of other fragments. Therefore, linear concatamers can be, and often are, produced which have complex and unpredictable structures. Also, some of the formed concatamers form enlarged circles that, later on, might interfere with the subsequent PCR amplification.

SUBSTITUTE SHEET

The formation of concatamers is clearly an unwanted side reaction, particularly as, under certain circumstances, concatamer production can be dominant, leading to a high excess of linear concatamers over circles.

5

In practice, each of the linear concatamers and enlarged circles containing the target sequence can undergo exponential PCR amplification because they contain the binding sites for the primers. This leads to the  
10 amplification of nonspecific products, which is clearly disadvantageous. IPCR is therefore critically dependent upon the formation of the correct circular nucleic acid fragment.

15 The size of the nucleic acid fragment is also a critical factor for the success of the IPCR technique. Ideally, the size of the initial linear excised target fragment and the corresponding circle must be within the range suitable for the polymerase chain reaction. This range is normally  
20 between 100 base pairs (bp) and several kilo base pairs (kb).

However, it is obvious that the actual size of the target fragment will very much depend upon the distribution pattern  
25 of the recognition sites for the restriction endonuclease used in the digestion of the original sample, for example of genomic DNA. In practice fragments greater than 2 or 3 kb are often formed. These large fragments cannot therefore be amplified by the IPCR technique.

30

Another disadvantage of the IPCR technique is that one achieves amplification of both of the flanking sequences. There may be situations when one needs to amplify just one of the flanking regions, particularly as there is now more  
35 and more an important need to do gene walking only in one direction.

It would therefore be advantageous to have a PCR technique that does not include a circularisation step but allows exponential amplification of a flanking sequence. A number of such schemes have been reported in the literature (e.g. 5 Shyamak and Ames [1989] *Gene* 84 1; Kalman et al [1990] *Biochem and Biophys Res Comm* 167 504; Roux et al [1990] *Biotechniques* 8 48 ; and Markham et al GB-A-2221909).

Each of these reported schemes employs the use of a special 10 cassette for ligation to each end of the sticky-ended nucleic acid fragments formed by digestion of a nucleic acid sample with a restriction enzyme. It is to be noted that each of the special cassettes ligates not only to the ends of the target fragment but also to the ends of the other 15 restriction fragments produced by the initial digestion of, for example, genomic DNA.

Each of the methods employs either a different cassette construction or a different sequence of reaction steps to 20 achieve a degree of selectivity during the amplification reaction.

In the Shyamak and Ames method ([1989] *Gene* 84 1), the cassette (which is called a vector) includes one of the 25 primer annealing regions. The excised target nucleic acid fragment includes the required other primer annealing region. Therefore, in theory, one could amplify, by exponential PCR, the nucleic acid region located between the cassette and the primer annealing region in the target 30 sequence. In this method, the ligated cassette-target fragment is exposed to both of the primers at the same time.

However, problems arise with this scheme because all of the excised fragments have ligated cassettes at each end which 35 contain one of the primer annealing regions. Thus, unwanted fragments will be amplified during the exponential PCR amplification of the target fragment because the cassette primer, once hybridised to the cassette, can be extended by

the polymerase from both ends of the unwanted fragments. This, of course, leads to the presence of a large number of unwanted nucleic acid fragments in the final mixture, making further analysis impossible. This reduces the efficiency  
5 and precision of the method.

The Shyamak and Ames method therefore has only a limited use and can only really be used for the amplification of "simple" DNA samples (e.g. from very simple prokaryotic  
10 organisms). The method cannot really be used for amplifying a fragment within a complex genomic DNA mixture (e.g. from a eukaryotic organism) because, in addition to the target fragment (usually present in only one or a few copies), millions of unwanted fragments having ligated cassettes will  
15 also be exponentially amplified.

Furthermore, since the cassette primer is only present in a limited quantity, most of the fragments including the target fragment will not be amplified because the primer will soon  
20 be exhausted. This is disadvantageous. Furthermore, if theoretically an excess of cassette primer is present, a mixture of millions of different fragments would be amplified. This is again clearly disadvantageous.

25 The reported Kalman and Roux methods (Kalman et al [1990] Biochem and Biophys Res Comm 167 504 and Roux et al [1990] Biotechniques 8 48) seek to overcome some of the above mentioned problems. However, there are similar problems associated with these methods (see below) which prevent them  
30 from being used for any given amplification problem. In each of these methods, synthetic oligonucleotide cassettes with sticky ends are used in the ligation reaction following the digestion of genomic DNA with a given restriction endonuclease. These cassettes ligate to the ends of all of  
35 the nucleic acid fragments (i.e. the target fragment and the unwanted fragments).

In the Kalman method, the cassette comprises two

complementary oligonucleotides that form a double-stranded piece of DNA having a sticky end. However, one oligonucleotide does not have a phosphate group at its 5'-end. Therefore, during the ligation reaction, only one of  
5 the oligonucleotides will covalently link to the excised nucleic acid fragments. The unligated oligonucleotide is then removed by selective ethanol precipitation in the presence of ammonium acetate. The ligated cassette thus becomes single stranded.

10

Next PCR amplification is carried out in the presence of both of the primers i.e. the primer complementary to a region of the target sequence and the primer complementary to a portion of the cassette.

15

The first cycle of the PCR amplification comprises only a linear extension of the primer annealed to the target sequence. In theory, it is only after this linear reaction step that the cassette primer can take part in the  
20 exponential PCR cycles by hybridising to the extension product. Therefore, and according to the reported method, unwanted nucleic acid fragments should not be amplified either in a linear or exponential fashion.

25 However, the proposed scheme does have some drawbacks. In particular, when the technique is used in the amplification of procaryotic genomic DNA, a large background of amplified fragments are observed, with only a slight excess production of the amplified target fragment.

30

Another drawback of the technique relates to the fact that the unligated oligonucleotide can never be quantitatively removed by precipitation. Even if the removal is 99% or 99.9% complete there will be enough unligated  
35 oligonucleotides associated with unwanted fragments to allow linear amplification of these unwanted fragments.

A third drawback stems from the fact that the incomplete DNA strand in the cassette serves as a primer and can therefore be elongated by polymerase in the presence of all four deoxynucleotide triphosphates to yield a complete  
5 double-stranded cassette at the ends of all excised fragments.

Taking into account each of these drawbacks, non-specific amplification will be a major part of the amplification step  
10 when both of the primers (i.e. one specific for the target sequence and one specific for the cassette) are present in the reaction mixture.

Therefore, the proposed Kalman method will only work  
15 successfully for very simple prokaryotic mixtures and, even then, high background levels of unwanted products will be observed.

The Roux method follows a pattern similar to the Kalman  
20 method, wherein a mechanism is introduced to produce, in theory, only one strand suitable for exponential PCR amplification.

In the Roux method, the cassette consists of two  
25 oligonucleotides of different lengths. The short strand is known as the tailed linker or incomplete strand. The longer strand, which is at least 15 to 20 nucleotides longer, is known as the anchor template or complete strand. The incomplete strand has a number of bases that are non-  
30 complementary to the complete strand. These mismatches should, in theory, prevent a filling-in reaction during the PCR amplification step by the action of the polymerase. Moreover, because there is not a region in the short strand that is complementary to the primer annealing region in the  
35 complete strand, there should only be one template produced by a linear PCR amplification step in the presence of the two primers (i.e. the primer complementary to the target sequence and the primer for the cassette). The ligated

cassette-target fragment itself is therefore not suitable for exponential PCR amplification but it is this linear extension product that is suitable for exponential PCR amplification. Thus, it is only after this first linear  
5 extension step that the second primer can hybridize to the extension product and create exponential amplification.

However, there are problems encountered with the Roux method. In particular, it is not uncommon for the  
10 incomplete strand to be filled-in to yield a strand that is complementary to the remainder of the complete strand. This filling-in reaction should be orders of magnitude less than in the Kalman method. However, in spite of the mismatches, this filling-in reaction will still produce templates of all  
15 of the unwanted fragments that are suitable for exponential PCR amplification.

In summary, the Kalman and the Roux methods do not overcome the problems experienced with the Shyamak and Ames method.  
20

The method of Markham *et al* (GB-A-2221909) apparently tries to overcome the problems associated with each of the above described methods. In brief, the Markham method, like the Kalman and Roux methods, includes the use of synthetic  
25 oligonucleotide cassettes (which are defined as vectorette cassettes) that are ligated to both ends of both the target fragments and the unwanted fragments present in the digested original sample, e.g. genomic DNA. The cassettes are designed to enhance the specificity of the PCR amplification  
30 step.

The Markham method has two variants which employ two types of oligonucleotide cassette. Each cassette is constructed so that, after ligation to the fragments, a cassette primer  
35 cannot be hybridised to the cassette itself. Instead, the cassette primer should, in theory, only hybridise to an extended primer product. This effect is achieved by constructing cassettes comprising two oligonucleotides that

are only partially complementary. The primary structures of both oligonucleotides have non-complementary middle portions which remain single-stranded.

- 5 In the first variant, the cassette comprises a short oligonucleotide and a long oligonucleotide. The short oligonucleotide is blocked at its 3'-end with either a dideoxynucleotide or another suitable nucleotide derivative. This prevents a filling-in reaction during the PCR  
10 amplification steps. The long oligonucleotide is at least 15 to 20 nucleotides longer at its 3'-end.

In the second variant, both of the oligonucleotides still possess a certain degree of non-complementarity but they are  
15 each more than 50 nucleotides long.

Therefore, it should only be after the primer complementary to the target sequence has been extended by one linear PCR cycle that the primer complementary to the cassette can  
20 hybridise to the extended DNA strand and thus take part in the PCR reaction to exponentially amplify the target fragment.

In the Markham method, as in the other earlier methods, the  
25 PCR amplification steps are usually carried out in the presence of both the two primers - i.e. a primer that is complementary to the target sequence and a primer that is complementary to the cassette portion.

- 30 In the first PCR cycle, only the primer complementary to the target sequence should be linearly extended. This should create a template suitable for exponential PCR because the specific primer that is complementary to the cassette can now hybridize to the first extended PCR product and can thus  
35 be extended itself.

In theory, therefore, by repeating the whole process several times (usually 20 to 40 cycles), only the ligated cassette-



target fragment should be amplified, whereas all unwanted restriction fragments should not be amplified. The amplification should, therefore, be highly specific because the special cassette design should exclude unwanted fragment  
5 amplification.

The Markham method can include the optional step of conducting two separate amplification reactions, namely a linear amplification followed by an exponential  
10 amplification. That is to say linear amplification for several cycles using only the primer complementary to the target sequence followed by the addition of the primer complementary to the cassette leading to exponential PCR amplification in the presence of the two primers.

15 One possible reason for splitting the original method into two separate amplifications might be that the amount of target fragment could be slightly enlarged (by a factor of 2 to 100 if up to 100 linear PCR cycles are carried out)  
20 before exponential PCR amplification starts. Therefore, this two stage process might only be necessary if one starts from a few copies of a target fragment or just one single target fragment (e.g. from egg or sperm cells).

25 However, even though it has been shown that exponential PCR will work on a single DNA molecule, it has been shown that in some cases too much DNA template often leads to a dramatic increase of non-specificity during exponential PCR amplification. Therefore, the second version of the Markham  
30 method will not significantly differ from the first version because exponential PCR will also take place in the original complex mixture of digested human genomic DNA which, in turn, will cause non-specificity during the PCR amplification steps.

35 The Markham method can also include the addition of S1 nuclease after several linear amplification steps have been carried out with the primer specific for the target

fragment. The S1 nuclease should in theory degrade all of the remaining single-stranded unwanted DNA fragments. This should increase the relative concentration of the ligated cassette-target fragments over the remaining background  
5 levels of the unwanted nucleic acids that are present. However, S1 is known to attack to a significant extent double-stranded DNA. Therefore, this approach is not always a practical solution to reduce the complexity of the reaction mixture before starting exponential PCR  
10 amplification.

It initially appears that the Markham method offers certain advantages over the earlier mentioned methods. In particular, the cassettes are designed to achieve a specific  
15 amplification of the target fragment following a linear PCR amplification step. Thus, in theory, even though the PCR amplifications are performed in the presence of both primers (i.e. the specific primer which hybridizes to the target sequence and the cassette primer which hybridizes to the  
20 extended product and not to the cassette itself), the first step should be a linear amplification of the target DNA fragment starting from the annealed specific primer. In all the other PCR cycles both primers take part and the amplification is therefore exponential.

25 However, there are problems associated with the Markham method. For example, if it is used for the amplification of genomic DNA, the PCR amplification steps have to be carried out in the presence of the whole genomic DNA mixture of  
30 restriction fragments. This genomic mixture will contain millions of different fragments, particularly in the case of complex genomes. These fragments will cause some degree of nonspecific amplification.

35 The non-specificity is due to the clustering of the restriction sites in genomic DNA of complex eukaryotic organisms. Therefore, many quite small restriction fragments will be present in the mixture which, certainly

after denaturing, will serve as primers during PCR and can cause a high degree of non-specificity.

Accordingly, and even though the Markham method initially  
5 appears technically easier to perform, it does lead to some degree of non-specificity with both complex genomic DNA mixtures (like human genomic DNA) and simpler DNA mixtures (e.g. procaryotic DNA).

10 Aside from the non-specificity of the Markham method, it does have some further drawbacks. For example, special oligonucleotide blocking groups (e.g. dideoxynucleotides at the 3'-end) or very long oligonucleotides (e.g. above 50 nucleotides in size) are necessary for the construction of  
15 the cassettes. These are expensive and difficult to produce, especially if many different cassettes are to be used.

Moreover, it is also not feasible to conduct simultaneous  
20 exponential amplification of different specific nucleotide fragments at the same time. Thus, as with the other earlier mentioned cassette-mediated PCR amplifications, this method is not suitable for a multiplexing process (see below).

25 The Markham method apparently includes an optional step of isolating the extended primer strand by use of a gel. However, it is to be noted that the isolation of the amplified target fragments is carried out after the exponential amplification stages. Also, it is a widely  
30 recognised fact that gel separation methods are not only laborious and time consuming, but they are really only effective for detecting and isolating large quantities of large sized strands. Small strands are difficult to separate from each other using gels.

35

The reported Markham method includes no other method for isolating the extended primer strands. In particular, there is no disclosure of an isolation step dependent upon the use

of a labelled primer.

An isolation procedure, using a labelled primer, has been reported in the literature (Hultman et al [1989] N.A.R. 17  
5 4937). However, in the Hultman method, the labelling is only carried out during the exponential PCR amplification of a target DNA strand. This labelling step is then followed by binding the labelled DNA extension products to a polymer, separating the strands and subsequently sequencing the  
10 polymer-bound DNA strands.

In order to overcome the problems associated with each of the above techniques and methods, we have developed a modified PCR method which allows the amplification of  
15 nucleic acid regions that flank a target sequence. In particular, the present method does not require the use of a circularisation step or the use of specially designed cassettes (e.g. cassettes having incomplete strands).

20 More importantly, the present method allows one to pick out specific target nucleic acid strands containing the target sequence from a reaction mixture containing many different strands. The specific picking out of the target strands ensures that exponential PCR amplification only occurs on  
25 the target strand.

Also, most of the walking and sequencing methods at present are based on DNA cloned into plasmids, phages, cosmids or yeast artificial chromosomes (YACs). Primary cloning and  
30 subcloning is very tedious and time-consuming. It has been shown that some regions of genomic DNA cannot be cloned at all or prove to be very difficult to clone. These portions of genomic DNA cannot therefore be analysed by presently available walking or sequencing techniques.

35

Also, other regions containing repetitive or other structures are difficult to clone in certain vector systems. Ordering of individual clones of one library by mapping or

fingerprinting as well as their sequencing could be greatly improved if one could easily walk and sequence through gaps of a given library by using an efficient, fast and specific *in vitro* amplification method starting from a target  
5 sequence. This would avoid the earlier necessary construction of different libraries from the DNA of one organism (or part of it) using different vector systems.

The present method is generally well suited for walking and  
10 sequencing along any piece of genomic DNA without the need for cloning. The present method can also be used for the detection of point mutations, deletions, and insertions within any genomic region of interest. This is especially advantageous for the detection of any modifications in the  
15 coding or non-coding regions of genes associated with genetic disorders, cellular disorders or infectious diseases. This gives one the potential to design specific diagnostics. It also allows the early determination of polymorphism for both alleles in many individuals, even down  
20 to the nucleotide level. Furthermore, the method has applications in the identification and sequencing of unclonable loci, the identification of YAC termini for physical mapping and the extension of partial cDNA clones. Also, since the method contains only straight forward  
25 biochemical reactions, it can easily be automated.

The present method thereby overcomes and avoids most of the aforementioned problems.

### 30 Summary of the Invention

According to a first aspect of the present invention there is provided a polynucleotide amplification method comprising the steps of:

35

- i. forming a ligation product by ligating a target fragment, having sticky ends and including a first primer annealing region of known sequence, with a

- cassette, having a sticky end complementary to one of the sticky ends of the target fragment, the cassette including a second primer annealing region of known sequence, such that in the ligation product the known second primer annealing region is remote from the first primer annealing region,
- 5
- ii. denaturing the ligation product,
- 10 iii. annealing only a first primer to the first primer annealing region, the first primer having attached thereto a separating label,
- iv. adding nucleotides to the bound primer by use of a polymerase enzyme to form an extension product,
- 15
- v. denaturing the ds nucleic acid extension product thus formed,
- 20 vi. optionally repeating steps 3 to 5, leading to linear amplification in the production of single stranded (ss) nucleic acid having the separating label attached thereto,
- 25 vii. isolating the prepared ss nucleic acid by binding the attached label to a support matrix having a group cooperatively bindable with the label, and
- viii. subjecting the isolated nucleic acid to exponential PCR amplification.
- 30

It is important to note that in the present method, a linear PCR amplification step is carried out first of all and independently from an exponential PCR amplification step.

35 This is achieved by introducing a label in the linear amplification step that allows purification of the target fragment on a solid support before the exponential PCR amplification.

Therefore, because only the labelled products (i.e. the primer extension products corresponding to the target fragment) will bind to the matrix, any unlabelled products  
5 (which will not bind to the matrix) can be washed away by using suitable solutions (e.g. buffers, alkaline solutions etc).

The isolation step allows the easy isolation of and rapid  
10 purification of only the labelled fragment (i.e. the strand corresponding to the target fragment). The labelled fragments can then be subjected to exponential PCR amplification in the absence of any of the unwanted fragments. This leads to an efficient, effective and highly  
15 specific method for isolating and amplifying a target fragment having regions of unknown sequence from complex genomic DNA mixtures of restriction fragments - i.e. it reduces a complex mixture of restriction fragments to a single fragment or multiplex of fragments.

20 Thus, in the present method, a linear PCR step is used to introduce a specific binding label into a strand that will be complementary to the ligated cassette target fragment. In principle, one linear PCR cycle with a labelled primer  
25 should be sufficient to separate the labelled fragment from a complex genomic mixture before exponential amplification takes place. The present method therefore reduces a complex DNA mixture (e.g. genomic DNA) to a very simple mixture. Furthermore, because the separation of the labelled strand  
30 is completely specific, only the correct template will be present for the exponential PCR amplification step.

Preferably, the target fragment is derived from a digestion of a sample of DNA. For example, the sample of DNA could be  
35 total genomic DNA of a prokaryotic or eukaryotic organism, mixtures of total genomic DNA from different organisms or different individuals, a DNA fragment cloned in a vector like a phage, cosmid, YAC or mixture of different cloned

DNA fragments.

Preferably, the sample of DNA is digested with a suitable restriction enzyme or with a combination of different  
5 restriction enzymes.

Preferably, steps 3 to 5 are repeated up to 100 times; advantageously up to 50 times.

- 10 To enhance further the specificity of the amplification, a more specialised exponential PCR amplification step could be conducted using a third specific PCR primer. This third primer would be a nested primer with respect to the first primer. In this case, the target fragment would have a  
15 third known primer annealing region distanced from the original first primer annealing region. Preferably, this third region is situated between the first primer annealing site and the second primer annealing site of the cassette.
- 20 If desired, the third primer would also have on it a separating label so that the amplified fragments can also readily be separated. This will be especially useful if there is a possibility that the first primer also bound to fragments other than the target fragment.

25

- The presence of the third primer annealing region in the target strand enables the exponential PCR amplification step to be conducted using a primer that is specifically complementary to either the first primer annealing region or  
30 the third primer annealing region. This is particularly advantageous because, by using the third primer annealing region, a further selection mechanism is introduced wherein PCR amplification only occurs with the target fragment and not with any unwanted fragments that may have become  
35 inadvertently bound to the support. It therefore introduces a means for ensuring that only the target fragment is exponentially amplified.



This would also be preferable if, for example, the first labelled specific primer hybridized to several different places within the genomic mixture, which was then subsequently extended from all of the annealed points. The  
5 nested primer ensures that only the target fragment is amplified.

The nested third primer may furthermore be used in a preferred reamplification step, which advantageously  
10 facilitates the isolation of the target fragment in adequate purity and quantity for direct sequencing. In the reamplification step, an aliquot of the exponentially amplified mixture is reamplified using the cassette primer and a nested primer specific for the target fragment.  
15 Optionally, the nested target fragment-specific primer may be the third primer. Alternatively, it may be a different primer, hybridising to a further known primer annealing region distanced from both the first and second known primer annealing regions. Optionally, this different primer may be  
20 a fourth primer, but it is envisaged that any number of further nested primers may be advantageously employed. Equally advantageously, each different primer will hybridise to a separate further known primer annealing region. Thus a fourth primer would hybridise to a fourth known primer  
25 annealing region on the target fragment.

Preferably, the aliquot may be taken from a dilution of the exponentially amplified mixture, for example a dilution between 1:1 and 1:100, most preferably a dilution of 1:50.  
30 Advantageously the aliquot will measure between 0.1 and 10 $\mu$ l, preferably 1 $\mu$ l.

The reamplification step provides added levels of specificity to the amplification reaction through the use of  
35 nested primers. In addition, the presence of matrix-bound DNA templates is eliminated, and thus the inefficiency of amplification associated with such templates is resolved.

Preferably, the separating label is attached to the 5' end of the first primer. The label can also be attached to one or more heterocyclic bases of the first primer.

- 5 Preferably, the separating label is a biotin label and the support matrix comprises streptavidin-coated beads. However, it is to be understood that other forms of labels and support matrices would suffice e.g. proteins and protein binding groups, antibodies and antibody binding groups, GCN4  
10 and other DNA binding proteins.

It is also to be noted that the matrix need not be in the form of a bead. The matrix can be in any appropriate form. For example, if the matrix was in the form of a rod, the  
15 target strands could be isolated simply by dipping the rod into the reaction mixture and then removing it. In this case, only the target strands will bind to the rod which can then, if necessary, be washed. This set up would be ideal for an automated machine.

20

In a further example, the matrix could represent the surface of a well of a microtiter dish so that target strands of many different samples could be easily isolated, simply by handling the whole microtiter dish. Again, this set up  
25 would be ideal for an automated machine.

Preferably, the present cassette comprises two complementary oligonucleotides having 3 or 4 nucleotide overhangs such that a sticky end is formed. The oligonucleotides can be in  
30 the range of 20 to 30 nucleotides long. These oligonucleotides will be easy to synthesise, particularly as it is known that specific primers labelled (e.g. with biotin) can easily be synthesized in a two-step procedure. However, it is to be appreciated that the current new  
35 methods allow incorporation of biotin or other labels during automated DNA synthesis. Moreover, there would be no need to purify the primers before using them as cassettes for the ligation reaction.

**SUBSTITUTE SHEET**

Advantageously, the cassette sequence contains a universal primer sequence.

- 5 Examples of appropriate cassettes that can be ligated to the target strands include:

an EcoRI cassette

10           5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'  
              3' d(GCAACATTTTGCCGGTCACGGTTCATTAA) 5'

a HindIII cassette

15           5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'  
              3' d(GCAACATTTTGCCGGTCACGGTTCATCGA) 5'

a BglIII cassette

20           5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'  
              3' d(GCAACATTTTGCCGGTCACGGTTCACTAG) 5'

an XbaI cassette

25           5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'  
              3' d(GCAACATTTTGCCGGTCACGGTTCAGATC) 5'

a PstI cassette

30           5' d(CGTTGTAAAACGGCCAGTGCCAAGTTGCA) 3'  
              3' d(GCAACATTTTGCCGGTCACGGTTCA) 5'

a HinfI cassette

35           5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'  
              3' d(GCAACATTTTGCCGGTCACGGTTCATNA) 5'

[wherein N = any of the four bases G,A,T,C].

Examples of some appropriate primer sequences include:

the M13 Sequencing Primer (-21)

5

5'd(TGT AAA ACG GCC AGT)3', and

the M13 Sequencing Primer (-40)

10

5'd(GTT TTC CCA GTC ACG AC)3'.

In the present method, the isolated extension product can be exponentially PCR amplified while still in the matrix support-bound state. In this case appropriate primers are repeatedly annealed to the primer annealing regions to form double stranded nucleic acids on the addition of nucleotides to the annealed primers by use of a polymerase enzyme. Then, on denaturing the formed double stranded nucleic acids, the extension products simply fall away from the polymer-bound template into the solution. These extension products could then serve as ordinary templates in further PCR cycles. The products are also easy to collect and do not need to be separated by means of gels etc.

25 The use of the polymer-bound DNA template for exponential PCR has the extra advantage that it can be kept for long periods under appropriate storage conditions. This allows one to return to the bound fragments at a later stage to conduct any further experiments or amplification steps.

30

If desired, the bound ligation product can be removed from the matrix before undergoing exponential PCR amplification.

Following their preparation, the PCR products can be sequenced by any of the existing dideoxy termination or chemical degradation techniques using radio- or fluorescently-labelled nucleotides or primers.

It is known that the probability of undertaking a successful walk along a piece of genomic DNA from a target sequence into an unknown region depends on the unknown distribution pattern of suitable restriction sites (within the PCR range). Since this distribution pattern is completely unknown, it is better to choose several different restriction endonucleases to digest genomic DNA. Generally between 2 and 30 different restriction enzymes should be used to find one which produces a target fragment having a size within the PCR range. In most cases 5 different restriction enzymes possessing 6 nucleotide long recognition sites are sufficient for this purpose, for example EcoRI, HindIII, XbaI, BglII and PstI. If these restriction enzymes do not produce suitable restriction fragments, an endonuclease recognising 4 or 5 nucleotides like HinfI could be used.

After the genomic DNA has been digested into a number of restriction fragments by, for example, a suitable combination of different restriction enzymes, a number of appropriate cassettes, each comprising specific sticky ends for the given restriction endonucleases, must then be separately ligated to the restriction fragments in a series of independent, parallel ligation reactions.

The present invention has the advantage that it is particularly well suited for undertaking a successful walk along a piece of genomic DNA from a known site into an unknown region.

In particular, the excision and ligation steps can be conducted in the same vessel. In this way, all of the multi-ligation products can be pooled (i.e. multiplexed). They can then be easily isolated at the same time or in turn (see below) before commencing the other steps of the present procedure (i.e. linear PCR amplification using labelled primers that are complementary to the target sequence followed by isolation and purification of the labelled

linear PCR products on a solid support matrix and, finally, exponential PCR amplification in presence of two primers). Therefore, using the present process, the number of reactions is reduced to a minimum.

5

In addition, the method is just as efficient for the simultaneous walking from different target nucleic acid fragments into a particular unknown region. This special type of "multiplexing" is performed using several specific oligonucleotide primers for the first linear PCR amplification step. Each primer will be complementary to a respective specific region within the different target fragments from which one wishes to walk into the unknown regions. Each of the primers can then be linearly extended at the same time and, also, in the same reaction tube.

These primers could carry the same separating labels. If the labels are the same, the different extension products can be isolated by use of the same support matrix. Each of the extension products can then be subjected to an exponential PCR amplification step using the cassette primer and a mixture of nested specific primers each carrying different separation labels.

After the exponential PCR amplification stage, a specific amplified fragment can be obtained and sequenced from the mixture by using a support matrix which allows specific isolation of the appropriate label. The linear extension product of interest is isolated from the mixture by its separation on a solid support matrix with an appropriate binding group thereon. It can then be subjected to exponential PCR.

It is to be appreciated that in such a multiplexing reaction the appropriate primers need not have the same attached label. This allows one to pick out specific extension products if and when desired.

Of course, however, it is clear that the differently labelled amplification products could be isolated at the same time simply by adding a mixture of support matrices (with appropriate binding groups thereon) to the reaction  
5 mixture at the same time.

Accordingly, large sections of unknown nucleic acids can be amplified, isolated and sequenced at the same time simply by picking out each of the fragments of interest. This is  
10 particularly advantageous if there is a need to walk and sequence from many different starting points on a genome into unknown directions. Also, if the initial target fragments are overlapping, the sequence of the larger fragment, from which the fragments were excised, can be  
15 determined at the same time.

In summation, the present invention is particularly useful for a method called "multiplexing", wherein a number of different target nucleic acid fragments can be produced at  
20 the same time by the addition of a number of different restriction enzymes. Cassettes with appropriate labels can then be annealed to the known regions of the target fragments. Each of the primers can then be extended at the same time and the extension products can then be isolated by  
25 use of the same support matrix.

The present method therefore allows simultaneous exponential amplification of different specific DNA fragments by a single PCR amplification step. Specific DNA fragments can  
30 be isolated and purified out of this mixture by using different solid support and affinity binding mechanisms. Also, different labels can be used in the initial linear amplification step.

35 The present invention is therefore particularly useful for preparing nucleic acids and allowing genomic walking along a large section of genomic nucleic acid, e.g. human nucleic acid.

In using the present method for genomic walking, the sequences obtained from the first primer (i.e. the primer that anneals to the target sequence) and the second primer  
5 (i.e. the primer that anneals to the cassette) confirm the overlapping sequences. This gives the necessary information to enable one to sequence the unknown region and to design new primers for the next genomic walking step.

10 According to a second aspect of the present invention there is provided the use of a ligation product, which ligation product comprises a target fragment ligated to a cassette, in a method of genomic walking in any direction along the genomic nucleic acid, wherein the target fragment includes  
15 a first primer annealing region of known sequence and has annealed thereto a primer which has attached thereto a separating label, and wherein the cassette includes a second primer annealing region of known sequence.

20 The target fragment can be a fragment excised from genomic nucleic acid.

According to a third aspect of the present invention, there is provided a first kit comprising:

25

(a) a sample of genomic nucleic acid;

(b) means for excising a target fragment of nucleic acid having a first primer annealing region of  
30 known sequence from the genomic nucleic acid;

(c) a cassette ligatable to the excised fragment and having a second primer annealing region of known sequence;

35

(d) a first primer, annealable to the first primer annealing region, having attached thereto a separating label;



- (e) a second primer annealable to the second primer annealing region;
- 5 (f) a support matrix having attached thereto a group cooperatively bindable to the separating label; and optionally
- (g) a third primer annealable to a third primer  
10 annealing region of known sequence upstream or downstream of the first primer annealing region; and further optionally
- (h) at least any one of the following: a buffer, a  
15 polymerase, a washing solution and a nucleotide solution.

Preferably, the sample of genomic nucleic acid is DNA and this can include the DNA from one or more different  
20 organisms or segments of genomic nucleic acids.

Advantageously, the excising means is a restriction enzyme or a group of different restriction enzymes, including, optionally, appropriate digestion buffers.

25

Preferably, the kit further comprises a number of cassettes ligatable to the excised fragment and having second primer annealing regions of known sequence.

30 Advantageously, the kit further comprises incubation buffer and/or a sample of T4 DNA ligase.

Preferably, the kit comprises a number of first primers, each annealable to the first primer annealing regions, and  
35 having attached thereto the same or a different separating label.

**SUBSTITUTE SHEET**

Advantageously, the kit comprises a number of second primers, each annealable to the second primer annealing regions.

- 5 The kit can include a number of support matrices, each having attached thereto the same or a different group that is cooperatively bindable to the separating labels.

Preferably, the kit further comprises a number of third  
10 primers, each annealable to the third primer annealing regions of known sequence that are situated on the target fragments, preferably located between the first and the second primer annealing regions. Optionally, the kit further comprises fourth and further primers hybridisable to  
15 fourth or further primer annealing regions of known sequence which are situated on the target fragment, between the first and second primer annealing regions.

The kit can also include at least any one of the following:  
20 a buffer for *in vitro* amplification, a deoxynucleotide triphosphate solution, a polymerase, light mineral oil, one or more washing solutions, and means to attach a separating label to a (or any) first oligonucleotide primer that is of specific interest to the user in application of this kit.

25 According to a fourth aspect of the present invention, there is provided a second kit comprising:

(a) a ligation product comprising a target fragment  
30 of genomic nucleic acid ligated to a cassette, wherein the fragment includes a first primer annealing region of known sequence and the cassette includes a second primer annealing region of known sequence;

35 (b) a first primer annealable to the first primer annealing region and having attached thereto a separating label;

- (c) a second primer annealable to the second primer annealing region;
- 5 (d) a support matrix having attached thereto a group cooperably bindable to the separating label; and optionally
- 10 (e) a third primer annealable to a third primer annealing region of known sequence upstream of the first primer annealing region; and further optionally
- 15 (f) at least any one of the following: a buffer, a polymerase, a washing solution and a nucleotide solution.

Preferably, the second kit has a number of ligation products.

20

Accordingly to a fifth aspect of the present invention, there is provided a method for extending cDNA clones using the PCR amplification method of the first aspect of the invention. A cDNA clone of which only a central portion has  
25 been sequenced can be extended to both the 5' and 3' termini using specific primers hybridising to known regions of the cDNA and general oligonucleotides which are hybridisable to the termini of the cDNA.

30 Preferably, the method of the fifth aspect of the invention comprises the following steps:

i) synthesising double-stranded cDNA with the first strand primed with a first primer which hybridises to the  
35 poly-A tail of the mRNA;

ii) linear amplification of an aliquot of the cDNA using a primer, the primer being hybridisable to only one strand

of a target cDNA, and having a separating label attached thereto;

iii) isolation of the labelled target cDNA extension  
5 product by binding the label to a support matrix having a group cooperably bindable with the label;

iv) tailing the 5' end of the PCR products resulting from the linear extension of a primer hybridising to the  
10 antisense strand with a dNTP; and

v) amplifying the unknown regions of the target cDNA between the known region and the 5' and/or the 3' terminus, by exponential amplification with an internal primer  
15 hybridising to the known region, and a general primer, which for the PCR product extended from the antisense strand-binding primer will comprise poly-dN where N is a base complementary to the dNTP used in step (iv) and for the PCR  
20 product extended from the sense strand-binding primer will comprise a primer complementary to or substantially identical to at least part of the first primer used in step (i).

The products of the PCR reaction may advantageously be  
25 visualised on an agarose gel, and directly sequenced if desired.

Preferably, the first primer hybridising to the poly-A tail may further comprise the sequence of the RACE primer  
30 (5'(ATCGATGGATCCGCGGCCGC(T)<sub>20</sub>)3'; M.A. Frohman et al., (1988), PNAS 85, p. 8998-9002). Advantageously, the linear amplification of the cDNA may proceed for between 1 and 100 cycles, preferably 50 cycles.

35 Preferably, the dNTP used in step (iv) is dGTP; and the poly-dN in step (v) is poly-dC.

Advantageously, the general primers used in step (v) for the extension product of the antisense strand-binding primer may comprise a restriction endonuclease cleavage site attached to poly-dN. Preferably, the cleavage site may be  
5 a BamH1 a cleavage site. Thus the primer may be AACGAT(C)<sub>15</sub>. For the extension product of the sense strand-binding primer, the general primer may be the RACE primer, ATCGATGGATCCGCGGCCGC.

10 The above technique has been found to be especially effective, more so than the original RACE technique (Frohman et al., Op. Cit.).

According to a sixth aspect of the invention there is  
15 provided a third kit, comprising:

- i) reagents suitable for the synthesis of cDNA from an RNA preparation and, optionally, an RNA preparation;
- 20 ii) a first strand priming primer hybridisable to the poly-A tail of an RNA species;
- iii) a primer attached to a separatable label, hybridisable to a primer annealing region on one strand of a target mRNA;
- 25 and, optionally, a further primer hybridisable to the alternative strand;
- iv) a support matrix having attached thereto a group cooperably bindable to the separatable label;
- 30 v) a dNTP;
- vi) a general primer hybridisable to a tail of the dNTPs in item (v);
- 35 vii) a 3' end general primer hybridisable with the first strand primer of item (ii); and optionally

vii) at least one nested internal primer hybridisable to one strand of the known region of the target mRNA.

Preferably, the kit contains primers hybridisable to both  
5 cDNA strands, in order to extend the cDNA in both directions.

The present invention has several advantages. Firstly, the present method allows the linear PCR amplification of a  
10 target strand within a complex genomic nucleic acid mixture followed by an effective isolation and purification of the extended product from this mixture. Thus, only target strands can be exponentially amplified in complete isolation from any other fragment.

15 The method therefore introduces a way of amplifying, with a very high degree of specificity, any target sequence from any complex genomic nucleic acid mixture. This is in direct contrast with all of the other known procedures, wherein the  
20 step of exponential amplification of the desired target sequence takes place in a very complex mixture of genomic nucleic acid fragments and/or genomic restriction fragments. This naturally leads to a certain degree of non-specificity.

25 Secondly, the present cassette constructions are very simple and cheap to manufacture. They consist of two complementary oligonucleotides having sticky 3 or 4 nucleotide overhangs. These oligonucleotides are in the 20 to 30 bp range. They can be easily synthesised and can be used for ligation  
30 without any purification.

Thirdly, ligation of the cassette to the target nucleic acid does not lead to any of the problems encountered with the known IPCR technique (e.g. formation of concatamers). This  
35 is because the cassettes have one sticky end (nucleotide overhand) and one blunt end (i.e. an end that does not have a nucleotide overhang). This prevents the formation of concatamers. Also, there is no need for a recircularisation

step because the cassette provides the required first primer annealing region.

Fourthly, the present method can be done with any preparation of a genomic nucleic acid fragment. In particular, the method is independent of the size and molecular weight of the fragment.

Fifthly, the PCR products can be easily sequenced from either end. The sequence obtained from the first primer (i.e. the primer that anneals to the primer annealing sequence in the target sequence) confirms the overlapping sequence. The sequence obtained from the second primer (i.e. the primer that anneals to the annealing sequence in the cassette) represents the last part of the unknown region and provides the necessary nucleotide information to design and synthesize newly labelled primers for a next genomic walking step.

In sequencing an amplified target nucleic acid fragment, one has the choice of applying any of the known dideoxy termination or degradation sequencing techniques. In doing so, one can use radio-labelled or fluorescently labelled primers or terminators. The present method is very well suited for, and is compatible with, any of the known sequencing methods including enzymatic and chemical fluorescent procedures which allow automated on-line detection of the nucleotide sequence during electrophoresis.

This is an important advantage because direct sequencing of PCR products is not always simple and straightforward. For instance, noncoding PCR templates from spacer or intron regions often have high GC or AT content and dideoxy sequencing techniques (even using Taq DNA polymerase) do not always allow one to determine sequences unambiguously. In these cases, chemical degradation techniques have to be applied.

Since the cassette contains the universal primer sequence (e.g. a M13 sequencing consensus sequence), commercially available primers including fluorescently labelled primers can be used to sequence the first 300 to 500 bp from the cassette into the unknown region. If the amplified target fragment is quite large in size it cannot be sequenced in one go from both ends (i.e. from the first and second primer annealing region) and "walking primers" have to be synthesized and used for the DNA sequencing.

10

As mentioned above, an important advantage of the present technique is that it, unlike the earlier methods, allows different samples to be pooled and processed simultaneously through the steps of linear amplification, isolation, purification and, finally, exponential amplification all at the same time and in the same reaction tube.

This "multiplex" strategy is based on digesting genomic nucleic acid with different restriction endonucleases, ligating appropriate cassettes to the produced restriction fragments, pooling the ligation products and processing them in a simultaneous fashion through all of the subsequent amplification, isolation and purification steps.

The efficiency of the method can still further be increased by using a number of different, specific primers that anneal to different primer annealing sequences on the target fragment. These primers can carry different separating labels for the exponential PCR step. The labelled amplified target fragments can then be isolated and purified from this mixture by using solid supports with appropriate binding groups. Multiplexing large numbers of different samples is not possible by using any of the earlier methods.

A further advantage of the new technique is that any or a number of restriction endonucleases can be used to cut the genomic DNA into a number of different fragments. The endonuclease can be chosen so that it cuts the genomic DNA



within a region of known sequence and within an adjacent region of unknown sequence. The cassettes can be designed so that they can anneal to the restriction endonuclease cut ends, as well as including the desired primer annealing regions.

The present method has an even further advantage that it eliminates or reduces the possibility of amplifying unwanted fragments. For example, if the specific primer exhibits non-specific hybridisation within a complex genomic DNA mixture (which can also happen with each of the aforementioned methods) one will get some nonspecific binding of certain DNA fragments on the solid support during separation. This will lead to a mixture of several different DNA fragments. However, these impurities will be present in quantities that are much lower than those described in each of the earlier methods. Moreover, the stringent separation conditions in the present method (e.g. high salt, alkaline conditions) will reduce the content of the unwanted DNA fragments even further before the exponential PCR amplification steps take place.

Using the new method, we have successfully walked along the nematode unc 31 gene contained on a yeast artificial chromosome (YAC) clone within the background of total yeast DNA (see discussion below and in particular figure 4).

We have also used the technique to walk along total human DNA from exon 51 of the Duchenne Muscular Dystrophy (DMD) gene into the adjacent intron and within this intron itself (see discussion below and in particular figure 5).

Furthermore, we have used the method including the reamplification step to extend the human microclone M54 in both directions, deriving new sequence data from the human CAM-L1 gene (see figure 6).

In all cases we have successfully elongated the known locus by several kilo base pairs (kb). Numerous control walks have also been made within known (already cloned and sequenced) parts of the nematode unc31 and the human DMD gene. In all cases, each round of cassette-mediated PCR walking was performed using multiple restriction enzymes (EcoRI, HindIII, XbaI, BglII, PstI and HinfI) and appropriate oligonucleotide cassettes for ligation.

#### 10 Brief Description of the Drawings

Three specific embodiments of the present invention will now be described and with reference to the accompanying drawings, in which:-

15

Figure 1 is a general scheme of one use of the present cassette-mediated PCR technique, namely the exponential amplification of an unknown nucleic acid sequence within a gene;

20

Figure 2 is a schematic diagram of the method of the invention comprising the step of reamplification of a sample of the amplified mixture using a nested third primer;

25 Figure 3 is a general scheme portraying the application of the method of the invention to cDNA clone extension;

Figure 4 is a representation of the result of a successful walk of a 1 kb nucleic acid sequence within the nematode unc 31 gene contained in a YAC clone within total yeast genomic DNA, following the general scheme of figure 1;

30 Figure 5 is a representation of the result of a successful walk of an about 600bp nucleic acid fragment within the Duchenne Muscular Dystrophy (DMD) gene extending the known nucleotide sequence of intron 50 by about 400 bp using total human genomic DNA, following the general scheme

of figure 1.

Figure 6 is a representation of the result of PCR walks extending human microclone M54 (Mackinnon et al., 1990 Am. J. Hum. Genet. 47, 181-186) by 700 bp in either direction, following the general scheme of figure 2.

#### Detailed Description of the Embodiments

10 In the general scheme of figure 1, the symbol R represents a restriction site, the symbol KL represents a known locus, the symbol UKL represents an unknown locus, the symbol ds NS represents a double stranded nucleotide sequence (such as genomic DNA), the symbol ss NS represents a single stranded  
15 nucleotide sequence, the symbol OC represents an oligo-cassette, the symbol B represents a biotin labelled specific primer, the symbol SB represents a streptavidin coated bead, the symbols OHC and OHR represent nucleotide overhangs, and the symbols P1 and P2 represent appropriate primers for  
20 exponential amplification by PCR.

Following the general scheme of figure 1, a target fragment of nucleic acid (ds NS) is first excised from a larger sequence at restriction sites (R) by the use of an  
25 appropriate restriction enzyme (see step 1).

Next, oligo-cassettes are ligated to the ends of the excised fragment (see step 2). Each cassette (OC) has a blunt end (E) and a nucleotide overhang (OHC) at its other end. The  
30 overhang (OHC) is complementary to the nucleotide overhang (OHR) that is created when the ds NS is cleaved at the restriction sites (R), thus enabling the cassette (OC) to be ligated to the ds NS.

35 In order to enable the required nucleic acid sequence to be isolated in a simple and straightforward manner, linear PCR amplification of the ligated cassette-target fragment is conducted using a specific biotin-labelled primer (B) (see

step 3).

The products of the linear PCR amplification step are then isolated by, for example, admixing the reaction mixture with  
5 streptavidin-coated magnetic beads (SB) (see step 4). The biotin-labelled PCR products (from step 3) selectively bind to the streptavidin-coated magnetic beads (SB). The products are thus easily separated.

10 Next, the separated biotin-labelled PCR products, while still bound to the coated magnetic beads, are denatured and then exponentially amplified by the PCR technique using the two appropriate primers (P1, P2) (see step 5).

15 It is important to realise that P2 need not have the same sequence as B. If it does not, then a further specificity is introduced into the scheme, wherein it is ensured that only the fragment of interest is subjected to PCR amplification. The double stranded nucleic acid products  
20 can then be sequenced by any standard sequencing technique (see step 6).

Once the products have been sequenced the neighbouring fragment of nucleic acid can be isolated and sequenced by  
25 repeating the above steps.

In Figure 2, a scheme similar to that in figure 1 is shown, but comprising an added reamplification step. In Figure 2 primer B is the nested primer, while primer A is the  
30 biotinylated first primer and primer C is the cassette-hybridising primer.

Linear amplification is allowed to proceed as for the method of Figure 1. The DNA sequences of interest are isolated on  
35 streptavidin-coated beads, and exponentially amplified using primers A and C. An aliquot of the product of the exponential amplification is then re-amplified using nested primer B and cassette primer C. The use of a nested primer,

**SUBSTITUTE SHEET**

which is complementary to a portion of the known sequence, adds a further level of specificity and improves the purity of the final product. This product can be sequenced directly without the need for cloning.

5

Figure 3 is a schematic representation of the application of the method of the invention to the extension of cDNA clones.

Step I comprises the synthesis of double stranded cDNA, the  
10 first strand having been primed with the RACE primer, 5'(ATCGATGGATCCGCGGCCGC(T)<sub>20</sub>)3', which hybridises to the poly-A tail of the mRNA. The region of the cDNA shown shaded black is the region whose sequence is known, while the unshaded regions represent unknown cDNA sequences.  
15 Primers A and B, which are biotin-labelled, and C and D are constructed complementary to regions of the known cDNA sequence as shown.

In step II, the cDNA is split into two aliquots and linearly  
20 amplified for 50 cycles using only the biotinylated primers A or B, as shown. Each primer hybridises to a different strand of the cDNA.

Step III involves the isolation of the biotinylated product  
25 on streptavidin beads.

Step IV is carried out only on the aliquot which has been amplified using the primer hybridised to the antisense strand of the cDNA. The 5' end of the resulting extension  
30 product is tailed with dGTP using terminal transferase.

This allows hybridisation of the 5' end of the strand with a poly-dC primer.

35 Step V is exponential amplification of the two cDNA populations using two primers. Primers C and D, which are nested within the biotinylated primers A and B and the terminal primers, are used together with a poly-dC terminal

primer for the antisense strand product and a RACE primer for the sense strand product. The products of the exponential amplification may be directly sequenced and cloned if necessary.

5

As stated above, figure 4 records the result of a successful walk of 1 Kb within the nematode unc 31 gene contained in a YAC clone and figure 5 records the result of a successful walk along a segment of total human DNA from exon 51 of the  
10 Duchenne Muscular Distrophy (DMD) gene into the adjacent intron and within this intron itself.

Figure 6 records the results of a bidirectional walk of 700 bp in each direction from human microclone M54.

15

Total human genomic DNA was digested in parallel with five different restriction enzymes: EcoRI (E), HindIII (H), XbaI (X), BglII (B) and PstI (P). Appropriate oligo-cassettes were ligated to the ends of all restriction fragments and  
20 PCR walking was carried out in parallel as described in figure 2 using two pairs of M54-specific oligonucleotides. After exponential amplification (step 4 in figure 2) an aliquot of each PCR product was analysed by electrophoresis using a 1% agarose gel.

25

Lanes P in panel A and B show two PCR products which extend the microclone M54 at both ends by approximately 700 bp towards two Pst sites. Lanes E and H in panel B also show other PCR products which are caused by hybridisation of the  
30 biotinylated primers to similar regions in the genome.

The experimental details will now be discussed.

#### Experimental Details

35

#### Figures 4 and 5

A sample of total genomic DNA was prepared. The total genomic DNA contained yeast genomic DNA (a recombinant YAC

with the unc31 gene), nematode genomic DNA and human genomic DNA. 250-500ng of the total genomic DNA was digested to completion in a 20 $\mu$ l solution with the six restriction enzymes EcoRI, HindIII, XbaI, BglII, PstI and HinfI which  
5 were then inactivated by heating (equivalent to step 1 in figure 1).

Half of the digested DNA (125-250ng) was ligated either to 5 pmol of the appropriate EcoRI, HindIII, XbaI, BglII or  
10 PstI oligo-cassettes or to 50 pmol of a HinfI oligo-cassette in a total volume of 20 $\mu$ l (i.e. 10 $\mu$ l of the appropriate digested genomic DNA, 2 $\mu$ l of 10 times T4 DNA ligase buffer containing 200 mM Tris/HCl pH 7.4, 100mM MgCl<sub>2</sub> and 100mM DTT, 2 $\mu$ l of 6mM rATP, 4 $\mu$ l water and 1 $\mu$ l of the appropriate  
15 oligo-cassette in a concentration 5 or 50 pmol/ $\mu$ l) (equivalent to step 2 in figure 1). Each of the double-stranded cassettes were 28 nucleotides long and had appropriate 4 nucleotide overhangs (EcoRI, HindIII, XbaI, BglII and PstI) or a 3 nucleotide (HinfI) overhang. The  
20 cassettes were prepared from crude oligonucleotides by mixing together equimolar amounts of both oligonucleotides representing the upper and lower strands of the cassette, heating the mixture for 10 min to 80°C and slowly cooling the solution to room temperature over a period of 30  
25 minutes. The reaction volume was then diluted with water to 100 $\mu$ l and heated for 10 min to 75°C to heat-inactivate T4 DNA ligase.

Next, 1 $\mu$ l (1.25 to 2.5ng) of the ligated product was  
30 amplified by linear PCR steps using a specific primer having a biotinylated 5'-end.

One of the walks along the nematode unc31 gene (figure 4) was performed using a 24 nucleotide long specific primer  
35 having the following sequence:

5' Biotin-d (CGT TTC GCC CGA TAC AAT AAC AAT) 3'.

In case of the elongation of the DMD intron 50 region within total human DNA (figure 5), a 24 nucleotide primer sequence was used having the following sequence:

5            5' Biotin-d (CAG CTG GGT TAT CAG AGG TGA GTG) 3'

(The addition of the labelled primers is equivalent to step 3 in figure 1.)

10    The linear PCR step was carried out in 1 x PCR buffer (10 mM Tris HCl, pH 8.3; 25°C; 50 mM KCl; 1.5mM MgCl<sub>2</sub>; 0.01% gelatin) (Cetus) with 250μM dNTP's, 0.5μM specific biotinylated primer and 2.5 units Taq polymerase (Cetus). The PCR rate was 50 cycles of 95°C for 0.5 minutes, 55°C for  
15    1 minute, and 72°C for 1 minute.

The biotin-labelled products were then isolated by mixing them with 25μl of washed streptavidin-coated beads (Dynal S.A.) (equivalent to step 4 in figure 1).

20    Following an incubation period of 15 minutes at room temperature, the beads were washed three times with 40μl of 1 M NaCl in TE buffer followed by three washes with 1 x TE buffer. After each washing stage, the supernatants were  
25    carefully removed.

The bead-bound DNA was then denatured by heat and subjected to exponential PCR amplification (equivalent to step 5 in figure 1). The conditions for exponential PCR amplification  
30    were similar to those for the linear PCR steps except that 35 cycles were carried out (instead of 50) and that two primers were present (instead of one), each of which was unbiotinylated.

35    In the case of walking along the nematode unc31 gene, the following cassettes were used: a 21 bp long universal primer complementary to a part of the second primer annealing region within the cassette having the sequence of:



5' d(CGT TGT AAA ACG GCC AGT) 3'

and either an unlabelled 24 nucleotide nematode unc31-specific primer complementary to a part of the first primer annealing region within the target fragment having the sequence:

5' d(CAG CTG GGT TAT CAG AGG TGA GTG) 3'

10

or a nested 24 nucleotide nematode unc31-specific primer complementary to a third primer annealing region within the target fragment having the sequence:

15 5' d(CTA CTC GAA TTG CTA TCC TAA TCT) 3'

In case of walking along the human DMD intron 50 the following cassettes were used, a 21 bp long universal primer complementary to a part of the second primer annealing region within the cassette having the sequence:

5' d(CGT TGT AAA ACG GCC AGT) 3'

and a nested DMD intron 50-specific unlabelled 24 nucleotide primer complementary to a third primer annealing region within the target sequence having the sequence:

5' d(GAG ACT CAC ACT GGA CAA CCA GTG) 3'.

30 The PCR amplification products were then separated on 1% low melting point (LMP) agarose, isolated and sequenced (equivalent to step 6).

The sequencing was performed by both a standard chemical degradation and by a dideoxy termination technique using radioactive and/or fluorescent labels.

**SUBSTITUTE SHEET**

Figure 6

The basic principle of the method is outlined in figure 2. It involves 5 steps including

1. Linear amplification of the desired DNA fragment using  
5 a biotinylated primer complementary to the known locus;
2. Isolation of biotinylated specific PCR products by separation using streptavidin-coated magnetic beads;
3. Exponential amplification of the isolated polymer-bound specific DNA fragments using two primers, a cassette-specific primer and the locus specific primer but without  
10 biotin (primer A, figure 2);
4. Re-amplification of the desired fragment from a small aliquot of the PCR mixture from step 3 using the cassette-specific primer and a second locus-specific primer which  
15 lies internal to the first locus-specific primer (primer B, figure 2);
5. Direct sequencing of the PCR product.

For ligation to restriction fragments possessing 5'  
20 overhangs we have used a double-stranded oligonucleotide cassette 28 nucleotides long having an additional 4 nucleotide overhang. The "upper" oligonucleotide is the same for all cassette constructions and contains the (-21) M13 primer sequence:

25 5' d(CGT TGT AAA ACG GCC AGT GCC AAG T)3'.

The "lower" oligonucleotides were synthesised for the restriction enzymes EcoRI, HindIII, XbaI, and BglI and their nucleotide sequences are:

5' d(AAT TAC TTG GCA CTG GCC GTC GTT TTA CAA CG) 3' EcoRI  
30 5' d(AGC TAC TTG GCA CTG GCC GTC CTT TAA CAA CG) 3' HindIII  
5' d(CTA GAC TTG GCA CTG GCC GTC GTT TAA CAA CG) 3' XbaI  
5' d(GAT CAC TTG GCA CTG GCC GTC GTT TTA CAA CG) 3' BglI  
(the overhang is underlined).

35 The lower oligonucleotides are not phosphorylated and therefore are not covalently bound to the restriction fragment during ligation.

For ligation to restriction fragments possessing 3' overhangs we have used a double-stranded oligonucleotide cassette 28 nucleotides long having a 4 nucleotide overhang. The "lower" oligonucleotide is the same for all cassette constructions and contains the complementary (-21) M13 primer sequence:

5' d(ACT TGG CAC TGG CCG TCG TTT TAC AAC G) 3'.

The "upper" oligonucleotide was synthesized for the restriction enzyme PstI and its sequence is:

10 5' d(CGT TGT AAA ACG ACG GCC AGT GCC AAG TTG CA) 3'.

EcoRI, HindIII, XbaI, BglII and PstI oligo-cassettes are prepared at 5 pmol/ $\mu$ l concentration by dissolving approximately 500 pmol of the upper and the respective lower oligonucleotide in 100  $\mu$ l water. The cassettes are heated  
15 to 80°C for 5 min and then slowly cooled down to RT over a period of 30 min before using in the ligation reactions. The cassettes are stored at -20°C and thawed on ice before use.

20 Ten  $\mu$ l of digested genomic DNA (approximately 100 ng) were combined with 2  $\mu$ l 10 x ligation buffer, 2  $\mu$ l 10 mM ATP, 1  $\mu$ l of the EcoRI, HindIII, XbaI, BglII, or PstI oligo-cassette (5 pmol/ $\mu$ l), 4  $\mu$  water and 1  $\mu$ l T4 DNA ligase (1 unit). The mixtures were incubated overnight at 16°C. 80  $\mu$ l of  
25 water were added and the mixture heated 10 min at 70°C to destroy the ligase. The cassette-ligated DNA was aliquoted and stored at -20°C. This represents a stock for over 100 walking reactions.

30 Linear PCR was performed in 20  $\mu$ l using 0.5 ml test tubes. The following items were added: 13.5  $\mu$ l water, 2.0  $\mu$ l 10 x PCR buffer (Cetus), 2.0  $\mu$ l 2.5 mM dNTP mix, 1.0  $\mu$ l cassette-ligated DNA (1 ng), 1.0  $\mu$ l 10  $\mu$ M biotin-labelled locus-specific primer, and 0.5  $\mu$ l Taq polymerase (2.5 units).  
35 After overlaying with light mineral oil the following cycles were performed: 95°C 90 s [95°C 30s, 55°C 60s, 72°C 60s] x 50, 72°C 180s. All cycles were performed using the maximum heating and cooling rates possible with the Techne PHC-1 or

PHC-2.

The biotinylated product was recovered and purified by the addition of 25  $\mu$ l beads directly to the PCR mixture. After  
5 incubation at room temperature with occasional mixing, the beads were sedimented with a strong magnet. The supernatant and the oil were removed and the beads washed 3 times with 40  $\mu$ l TE/0.1 M NaCl and 3 times with TE. The beads were finally resuspended in 4.5  $\mu$ l H<sub>2</sub>O.

10

Two primers were used to amplify the single-stranded template bound to the beads: the locus-specific primer from step 1 of the method but without biotin (primer A, figure 2) and the cassette-specific primer (primer C, figure 2) with  
15 the following sequence: d(TGT AAA ACG ACG GCC AGT GCC) containing the M13 universal forward primer sequence. Exponential PCR is performed in 20  $\mu$ l comprising 9.0  $\mu$ l water, 2.0  $\mu$ l 10 x PCR buffer, 2.0  $\mu$ l 2.5 mM dNTP mix 1.0  $\mu$ l 10  $\mu$ M cassette-specific primer, 1.0  $\mu$ l 10  $\mu$ M locus-specific  
20 primer, 4.5  $\mu$ l bead bound DNA template, 0.5  $\mu$ l Taq polymerase. After overlaying with light mineral oil the following cycles were performed: 95°C 90s, [95°C 30s, 55°C 60s, 72°C 60s] x 35, 72°C 180s.

25 In order to isolate the extended product with sufficient purity and in adequate quantity for direct sequencing, a fraction (1  $\mu$ l from a 1 in 50 dilution) of the first exponential amplification was reamplified using a nested locus specific primer and the cassette primer (primers B and  
30 C, figure 2). Reaction conditions were similar to those described in the exponential amplification step.

Prior to direct sequencing, the PCR product was purified by gel electrophoresis using LMP agarose, in order to remove  
35 excess nucleotides and primers, as well as minor DNA contaminants. The DNA can be recovered using a Qiagen gel extraction kit, a Gene Clean II or Mermaid kit (Bio 101). Sequencing of PCR products was performed by linear

amplification sequencing using Taq polymerase and the fluorescnet dye terminators from Applied Biosystems (Taq Dye Deoxy <sup>TM</sup> Terminator Cycle Sequencing Kit). After cycling the fluorescent products were purified by G50 spin columns, lyophilized and loaded into a single lane of a 373A sequencer (Applied Biosystems). Reliable sequence information can be obtained in most cases from both ends of the PCR product using the M13 (-21) sequencing primer and the nested locus-specific primer. Walks along the PCR product were performed easily using 20mer synthetic oligonucleotides. The ends of each PCR product, including the primer and cassette sequence, were determined by solid-phase chemical degradation using radio-labelled, reamplified PCR products (16).

15

cDNA was synthesised from 0.1 µg of human fetal brain mRNA using AMV reverse transcriptase (Anglican Biotechnology, Colchester UK) under standard conditions but using the RACE-oligo-dT primer:

20 5'ATCGATGGATCCGCGCCGCTTTTTTTTTTTTTTTTTTTT3'.

The reverse transcriptase was heat-inactivated and the cDNA diluted to 100µl. A portion of this single stranded cDNA (1 µl) was then used for PCR.

25 For PCR amplification in a total of 50 µl the following reaction constituents were combined: 32.5 µl water, 5µl 10 x PCR buffer (Cetus), 5 µl 2.5 mM dNTP mix, 2.5 µl of each primer 5'(TTT GTC GAC) and primer 5' d(TTT GTC GAC) (the underlined portions representing a tail containing a Sali site), 2 µl human fetal brain cDNA (2ng), and 1 µl (5 units) Taq polymerase (Cetus). The mixture was overlaid with 40 µl of light mineral oil. Standard PCR cycles were 95°C 60s, [95°C 30s, 55°C 30s, 72°C 180s] x 40.5 µl of the PCR product were sized on a 1% agarose gel. The remaining 45 µl were  
35 passed through a Strategene PrimeErase<sup>TM</sup> column according to the manufacturer's recommendation in order to remove excess nucleotides, primers and polymerase, digested with Sali and subsequently cloned into pUC18. Recombinant colonies having

the correct insert size of about 2.4 kb were identified using either a shortened miniprep procedure (17) or by PCR amplification in a microtiter dish using universal forward and reverse primer and DNA prepared from a 200 ml culture according to the Qiagen protocol. Sequencing of double-stranded DNA was performed by linear amplification using the ABI dye terminators and a 373A sequencer. The insert was completely sequenced on both strands by adopting a walking protocol using 20-mer synthetic primers. Individual reads were between 350 and 400 bp.

### Experimental Results

The results are recorded in figures 4, 5 and 6.

15

Figure 4 shows a successful 1 kb walk along the nematode unc31 gene within total yeast DNA. In particular, lane 1 of figure 4 shows the predicted 1 kb band resulting from the exponential amplification of the target DNA between the nematode unc31-specific primer:

5' d(CAG CTG GGT TAT CAG AGG TGA GTG) 3'

and a HindIII site to which the cassette was ligated.

25

Lane 2 shows the result of exponential amplification with a nested nematode unc31-specific primer:

5' d(CTA CTC GAA TTG CTA TCC TAA TCT) 3'.

30

Lane 3 shows a range of molecular weight markers (123 base pair ladder, Bethesda Research Laboratories).

The PCR product from the exponential PCR amplification step with the cassette specific primer and the nested nematode unc31-specific primer (lane 2) was subjected to DNA sequencing. The results of which show the expected nucleotide sequence confirming the walk.

The short molecular weight bands in lane 1 and 2 represent side products which are often observed during PCR and are due to some side reaction with the applied primers. These 5 bands are not caused by the walking method itself.

Figure 5 shows a successful 600 bp walk along the human DMD gene within total human genomic DNA.

10 In particular, lane 1 shows a 600bp walk along the DMD intron 50 resulting from exponential amplification with a nested human DMD intron 50-specific primer:

5' d(GAG ACT CAC ACT GGA CAA CCA GTG)

15

towards a PstI site to which the cassette was ligated. This walk elongated the known portion of the DMD intron 50 by 400 nucleotides.

20 Lane 2 shows a range of molecular weight markers (123 base pair ladder, Bethesda Research Laboratories).

The PCR product from the exponential PCR amplification step with the cassette specific primer and the nested human DMD 25 intron 50-specific primer (lane 1) was subjected to DNA sequencing. The results of which show the expected overlapping nucleotide sequence confirming the walk. Around 400 bp were obtained from the end of the cassette and this represented new DMD intron 50 sequence. This was used to 30 synthesise a new biotinylated specific primer for the next cycle.

From the results, it is seen that the sequences obtained from the present oligo-cassette mediated PCR technique are 35 in agreement with those known for the nematode unc31 gene and those that are partially known for the human DMD gene.

Figure 6 shows the extended sequence derived from human microclone M54 of the human CAM-L1 gene.

A single PCR product was produced after the reamplification step 4 (figure 6, panel A, lane P). This suggests that a PstI site is located approximately 700 bp upstream of M54. The failure to amplify a product after digestion with the other enzymes suggest that these cut too far away from the locus. The use of a second primer set, directed downstream of M54 resulted in three different PCR products as can be seen from the agarose gel (figure 6 panel B, lanes E, H and P) and would suggest that three restriction sites (EcoRI, HindIII and PstI) might be located downstream from M54. Direct sequencing of the ends of each fragment by solid-phase chemical degradation revealed that only the two 700 bp long PCR products containing a PstI site at their ends are the real extension products of the microclone M54 because they showed the correct overlapping sequence between the internal primer and the ends of the microclone.

The 800 bp long HindIII-PCR product (figure 6, panel B, lane H) is not a extension of the M54 microclone. It was identified as an L1 repeat. The sequence of the HindIII-product matches the 11 kb long human L1 repeat located in the intergenic region of the epsilon and gamma globin gene between nucleotide positions 7744 and 8544 and shows about 80% homology. The human L1 repeat has a predicted HindIII site at position 7744-7750. However, the sequence of the microclone M54 is not contained within the above mentioned L1 repeat nor does the primer set used for walking show any significant match. Hybridisation experiments with M54, on the other hand, show that the human genome has more copies of this microclone. Presumably, the HindIII walk represents an extension from such an M54-like sequence into an adjacent L1 repeat.

The EcoRI PCR product (figure 6, panel B, lane E) did not show any significant matches with any sequences in the



database.

It will of course be understood that the present invention has been described above by way of example only and that  
5 modifications and variations can be made by the skilled person without departing from the scope of the invention.

**SUBSTITUTE SHEET**

## CLAIMS

1. A polynucleotide amplification method comprising the steps of
- 5
- i. forming a ligation product by ligating a target fragment, having sticky ends and including a first primer annealing region of known sequence, with a cassette, having a sticky end complementary to one of
- 10 the sticky ends of the target fragment, the cassette including a second primer annealing region of known sequence, such that in the ligation product the known second primer annealing region is remote from the first primer annealing region,
- 15
- ii. denaturing the ligation product,
- iii. annealing only a first primer to the first primer annealing region, the first primer having
- 20 attached thereto a separating label,
- iv. adding nucleotides to the bound primer by use of a polymerase enzyme to form a double-stranded nucleic acid extension product,
- 25
- v. denaturing the formed double stranded nucleic acid extension product,
- vi. optionally repeating steps 3 to 5, leading to
- 30 linear amplification in the production of single stranded nucleic acid having the separating label attached thereto,
- vii. isolating the prepared nucleic acid by binding
- 35 the attached label to a support matrix having a group cooperatively bindable with the label, and
- viii. subjecting the isolated nucleic acid to exponential

## PCR amplification.

2. A method according to claim 1 wherein steps iii to iv are repeated up to 100 times.
- 5 3. A method according to claim 2 wherein the steps are repeated about 50 times.
4. A method according to any preceding claim wherein the  
10 exponential PCR amplification step is conducted using a nested PCR primer that anneals to a third primer annealing region distanced away from the original first primer annealing region.
- 15 5. A method according to claim 4 wherein the third known primer annealing region is situated between the first primer annealing site and the second primer annealing site of the cassette.
- 20 6. A method according to any preceding claim wherein the separating label is attached to the 5' end of the first primer.
7. A method according to any preceding claim wherein the  
25 separating label is a biotin label and the support matrix comprises a streptavidin coated matrix.
8. A method according to claim 7 wherein the matrix is in the form of a bead or rod.
- 30 9. A method according to claim 7 wherein the matrix is the surface of a well of a microtiter dish.
10. A method according to any preceding claim wherein the  
35 cassettes comprise two complementary oligonucleotides that are each in the range of 20 to 30 nucleotides long.
11. A method according to claim 10 wherein the cassette

sequence contains a universal primer sequence.

12. A method according to claim 11 wherein the primer sequence is

5

5'd(TGT AAA ACG GCC AGT) 3', or

5'd(GTT TTC CCA GTC ACG AC) 3'.

10 13. A method according to any one of claims 10 to 12 wherein the cassette has the sequence of

5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'

3' d(GCAACATTTTGCCGGTCACGGTTCATTAA) 5',

15

5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'

3' d(GCAACATTTTGCCGGTACACGGTTCATCGA) 5',

5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'

20

3' d(GCAACATTTTGCCGGTCACGGTTCACTAG) 5',

5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'

3' d(GCAACATTTTGCCGGTCACGGTTCAGATC) 5',

25

5' d(CGTTGTAAAACGGCCAGTGCCAAGTTGCA) 3'

3' d(GCAACATTTTGCCGGTCACGGTTCA) 5 ', or

5' d(CGTTGTAAAACGGCCAGTGCCAAGT) 3'

3' d(GCAACATTTTGCCGGTCACGGTTCATNA) 5',

30

wherein N = any of the four bases G,A,T,C.

14. A method according to any preceding claim wherein the  
35 denatured ligation product is exponentially PCR amplified while still in the matrix support-bound state.

15. A method according to any preceding claim wherein the

**SUBSTITUTE SHEET**

PCR products from step viii are sequenced by any of the existing dideoxy termination or chemical degradation techniques using radio- or fluorescently-labelled nucleotides or primers.

5

16. A method according to any preceding claim wherein a number of ligation products are formed in step i and a number of first primers are added in step iii and a number of second primers are added in step viii.

10

17. A method according to claim 16 wherein each of the first primers has a different separating label.

18. A method according to claim 17 wherein a number of support matrices are added, each matrix having attached thereto a respective different group.

19. A method according to any preceding claim wherein the target ds nucleic acid fragment is derived from a digestion of a sample of DNA.

20. A method according to claim 19 wherein the sample of DNA is genomic DNA.

21. The use of a ligation product, which product comprises a target fragment ligated to a cassette, in a method of genomic walking in any direction along the genomic nucleic acid, wherein the target fragment includes a first primer annealing region of known sequence and having annealed thereto a primer which has attached thereto a separating label, and wherein the cassette includes a second primer annealing region of known sequence.

22. A first kit comprising:

35

(a) a sample of genomic nucleic acid;

(b) means for excising a target fragment of nucleic

acid having a first primer annealing region of known sequence from the genomic nucleic acid;

- 5 (c) a cassette ligatable to the excised fragment and having a region corresponding to a second primer annealing region of known sequence;
- 10 (d) a first primer, annealable to the first primer annealing region, having attached thereto a separating label;
- (e) a second primer annealable to the second primer annealing region;
- 15 (f) a support matrix having attached thereto a group cooperatively bindable to the separating label; and optionally
- 20 (g) a third primer annealable to a third primer annealing region of known sequence upstream or downstream of the first primer annealing region; and further optionally
- 25 (h) at least any one of the following: a buffer, a polymerase, a washing solution and a nucleotide solution.
- 30 23. A kit according to claim 22 wherein the sample of genomic nucleic acid is DNA.
24. A kit according to claim 23 wherein the DNA is from one or more different organisms or segments of genomic nucleic acids
- 35 25. A kit according to anyone of claims 22 to 24 wherein the excising means is a restriction enzyme.
26. A kit according to claim 25 wherein a number of

restriction enzymes are included.

27. A kit according to anyone of claims 22 to 26 wherein  
a number of cassettes ligatable to the excised fragment(s)  
5 are provided.

28. A kit according to anyone of claims 22 to 27 wherein  
the kit comprises a number of first primers, each annealable  
to the first primer annealing regions, and having attached  
10 thereto the same or a different separating label.

29. A kit according to anyone of claims 22 to 28 wherein  
the kit comprises a number of second primers, each  
annealable to the second primer annealing regions.

15

30. A kit according to anyone of claims 22 to 29 wherein  
the kit can include a number of support matrices, each  
having attached thereto the same or different group that is  
cooperatively bindable to the separating labels.

20

31. A kit according to anyone of claims 22 to 30 wherein  
the kit further comprises a number of third primers, each  
being annealable to the third primer annealing regions of  
known sequences that are situated on the target fragments.

25

32. A kit according to anyone of claims 22 to 31 wherein  
the kit includes anyone of an amount of incubation buffer,  
a sample of T4 DNA ligase, a buffer for in vitro  
amplification, a deoxynucleotide triphosphate solution, a  
30 polymerase, light mineral oil, one or more washing  
solutions, and means to attach a separating label to a (or  
any) first oligonucleotide primer.

33. A second kit comprising:

35

(a) a ligation product comprising a fragment of  
target genomic nucleic acid ligated to a  
cassette, wherein the target fragment includes a

first primer annealing region of known sequence  
and the cassette includes a second primer  
annealing region of known sequence;

- 5 (b) a first primer, annealable to the first primer  
annealing region, having attached thereto a  
separating label;
- (c) a second primer annealable to the second primer  
10 annealing region;
- (d) a support matrix having attached thereto a group  
cooperably bindable to the separating label; and  
optionally
- 15 (e) a third primer annealable to a third primer  
annealing region of known sequence upstream of  
the first primer annealing region; and further  
optionally
- 20 (f) at least any one of the following: a buffer, a  
polymerase, a washing solution and a nucleotide  
solution.
- 25 34. A kit according to claim 33 wherein the second kit has  
a number of ligation products.
35. A method for extending cDNA clones comprising a PCT  
amplification method as claimed in claim 1.
- 30 36. A method according to claim 35, comprising the steps  
of
- i) synthesising double-stranded cDNA with the first  
strand primed with a first primer which hybridises to the  
35 poly-A tail of the mRNA;
- ii) linear amplification of an aliquot of the cDNA using  
a primer, the primer being hybridisable to only one strand



of a target cDNA, and having a separating label attached thereto;

iii) isolation of the labelled target cDNA extension  
5 product by binding the label to a support matrix having a group cooperably bindable with the label;

iv) tailing the 5' end of the PCR products resulting from the linear extension of a primer hybridising to the  
10 antisense strand with a dNTP; and

v) amplifying the unknown regions of the target cDNA between the known region and the 5' and/or the 3' terminus, by exponential amplification with an internal primer  
15 hybridising to the known region, and a general primer, which for the PCR product extended from the antisense strand-binding primer will comprise poly-dN where N is a base complementary to the dNTP used in step (iv) and for the PCR product extended from the sense strand-binding primer will  
20 comprise a primer complementary to or substantially identical to at least part of the first primer used in step (i).

37. A method according to claim 35 or claim 36, further  
25 comprising direct sequencing of the extension product.

38. A method according to any one of the claims 35 to 37, where the first primer comprises oligo-dT and the RACE primer.  
30

39. A method according to any one of claims 35 to 38, wherein the first primer is  
5' (ATCGATGGATCCGCGGCCGC(T)<sub>20</sub>) 3'.

35 40. A method according to any one of claims 35 to 39, wherein step (IV) is accomplished using a terminal transferase.

41. A method according to any one of claims 35 to 40, wherein the dNTP is dGTP, and the poly-dN is poly-dC.

42. A method according to any one of claims 35 to 41,  
5 wherein the poly-dC general primer further comprises a sequence encoding a restriction endonuclease deavage site.

43. A method according to claim 42 where the primer is AACGAT(C)<sub>15</sub>.

10

44. A method according to any one of claims 38 to 43, wherein the general primer for the extension product of the sense-strand-binding primer is the RACE primer, ATCGATGGATCCGCGGCCGC.

15

45. A kit comprising:

(i) reagents suitable for the synthesis of cDNA from an RNA preparation and, optionally, an RNA preparation;

20 ii) a first strand priming primer hybridisable to the poly-A tail of an RNA species;

iii) a primer attached to a separatable label, hybridisable to a primer annealing region on one strand of a target mRNA;  
25 and, optionally, a further primer hybridisable to the alternative strand;

iv) a support matrix having attached thereto a group cooperably bindable to the separatable label;

30

v) a dNTP;

vi) a general primer hybridisable to a tail of the dNTPs in item (v);

35

vii) a 3' end general primer hybridisable with the first strand primer of item (ii); and optionally

vii) at least one nested internal primer hybridisable to one strand of the known region of the target mRNA.

46. A kit according to claim 45, comprising primers  
5 hybridisable to both strands such that the cDNA can be extended in both directions.

47. A kit according to claim 46 or claim 45, further comprising a terminal transferase.

FIG. 1

1/4

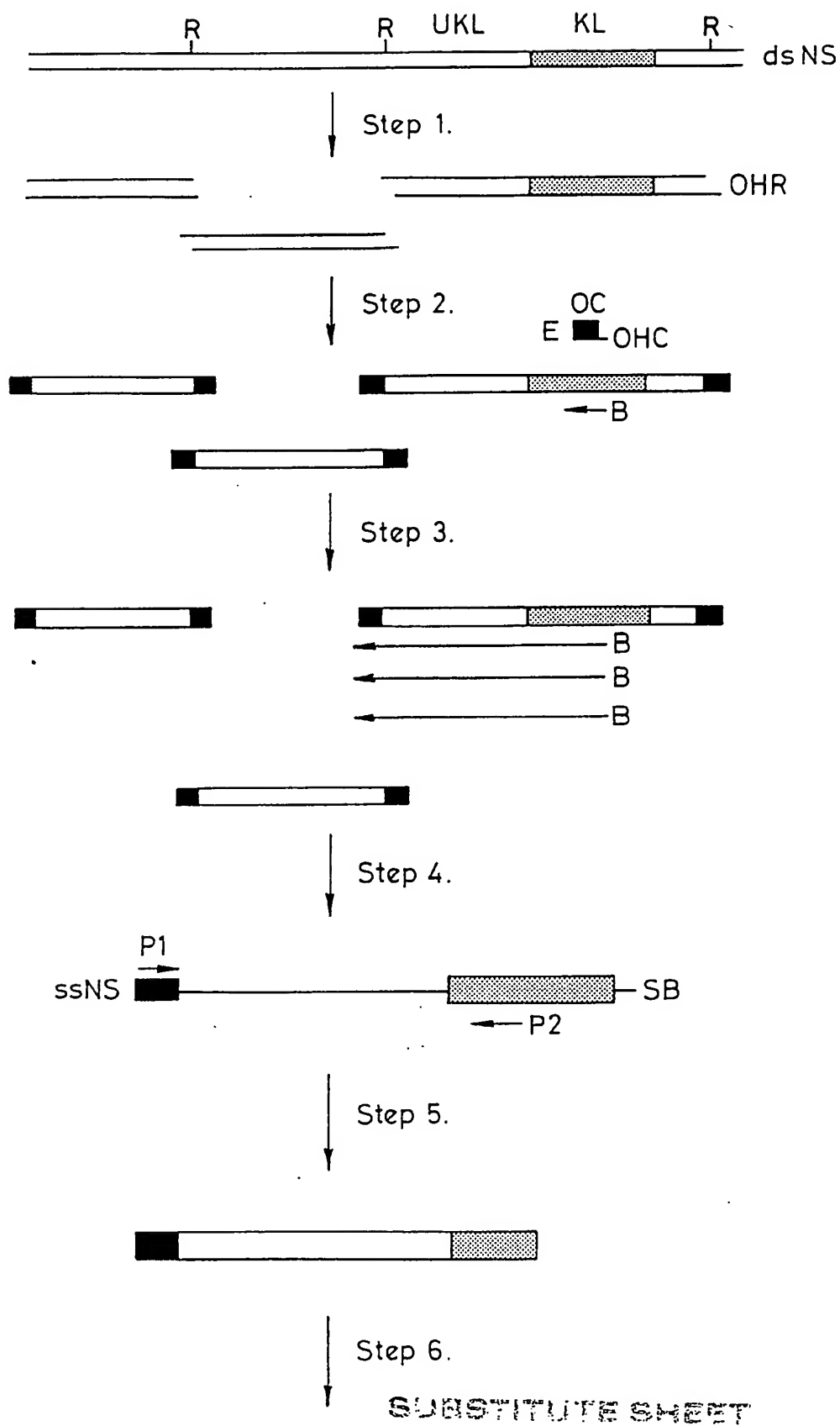
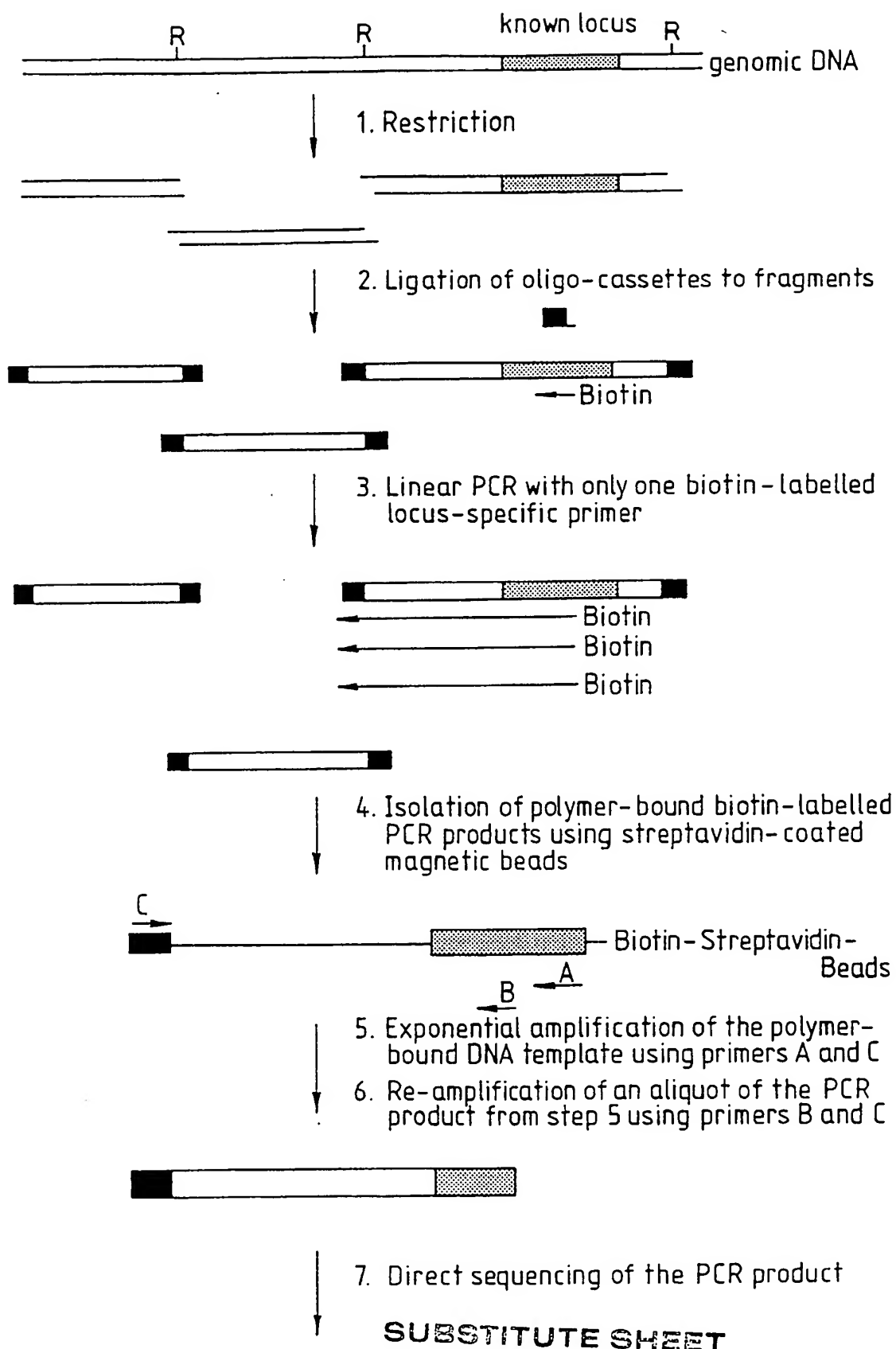
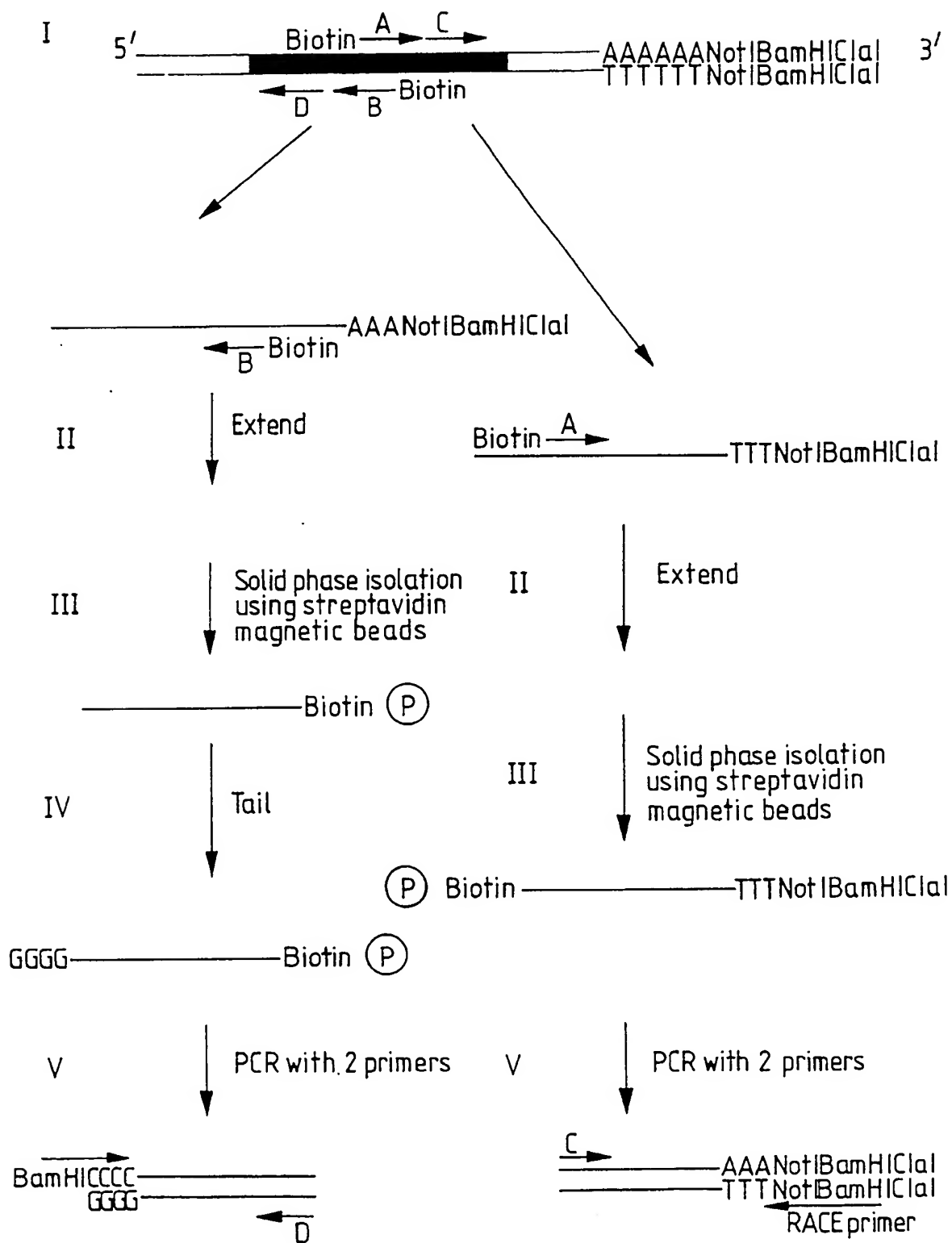


FIG. 2

2/4



3/4



SUBSTITUTE SHEET

4/4

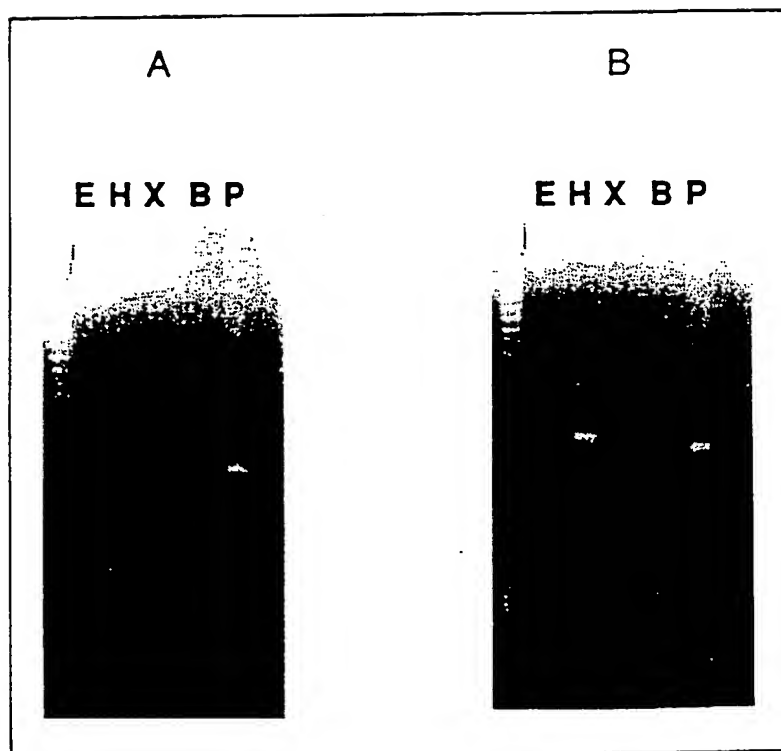


FIG. 4



FIG. 5

FIG. 6



# INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 91/00803

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (if several classification symbols apply, indicate all) <sup>6</sup> According to International Patent Classification (IPC) or to both National Classification and IPC IPC <sup>5</sup> : C 12 Q 1/68, C 07 H 21/00, //C 12P 19/34, C 12 N 15/10																				
<b>II. FIELDS SEARCHED</b> <div style="text-align: right; margin-right: 100px;">Minimum Documentation Searched <sup>7</sup></div> <table style="width: 100%; border-collapse: collapse;"> <tr> <th style="width: 30%; border-bottom: 1px solid black;">Classification System</th> <th style="border-bottom: 1px solid black;">Classification Symbols</th> </tr> <tr> <td style="border: 1px solid black; padding: 5px;">IPC<sup>5</sup></td> <td style="border: 1px solid black; padding: 5px;">C 12 Q</td> </tr> </table> <div style="text-align: center; margin-top: 5px;">Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched <sup>8</sup></div>			Classification System	Classification Symbols	IPC <sup>5</sup>	C 12 Q														
Classification System	Classification Symbols																			
IPC <sup>5</sup>	C 12 Q																			
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>9</sup></b> <table style="width: 100%; border-collapse: collapse;"> <tr> <th style="width: 10%; border-bottom: 1px solid black;">Category <sup>9</sup></th> <th style="width: 70%; border-bottom: 1px solid black;">Citation of Document, <sup>11</sup> with Indication, where appropriate, of the relevant passages <sup>12</sup></th> <th style="width: 20%; border-bottom: 1px solid black;">Relevant to Claim No. <sup>13</sup></th> </tr> <tr> <td style="text-align: center; vertical-align: top; padding: 5px;">A</td> <td style="padding: 5px;">EP, A, 0356021 (IMPERIAL CHEMICAL INDUST.) 28 February 1990 see the whole document cited in the application --</td> <td style="text-align: center; vertical-align: top; padding: 5px;">1-34</td> </tr> <tr> <td style="text-align: center; vertical-align: top; padding: 5px;">A</td> <td style="padding: 5px;">WO, A, 90/01065 (GENELABS INC.) 8 February 1990 see the whole document --</td> <td style="text-align: center; vertical-align: top; padding: 5px;">35-47</td> </tr> <tr> <td style="text-align: center; vertical-align: top; padding: 5px;">A</td> <td style="padding: 5px;">EP, A, 0297379 (MOLECULAR DIAGNOSTICS) 4 January 1989 --</td> <td></td> </tr> <tr> <td style="text-align: center; vertical-align: top; padding: 5px;">A</td> <td style="padding: 5px;">Biotechniques, volume 8, January 1990, K.H. Roux et al.: "A strategy for single site PCR amplification of dsDNA: Priming digested cloned or genomic DNA from an anchor-modified restriction site and a short internal sequence", pages 48-57 cited in the application --</td> <td></td> </tr> <tr> <td style="text-align: center; vertical-align: top; padding: 5px;">A</td> <td style="padding: 5px;">Gene, volume 84, 1989, Elsevier Science Publishers B.V. ./.</td> <td></td> </tr> </table>			Category <sup>9</sup>	Citation of Document, <sup>11</sup> with Indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>	A	EP, A, 0356021 (IMPERIAL CHEMICAL INDUST.) 28 February 1990 see the whole document cited in the application --	1-34	A	WO, A, 90/01065 (GENELABS INC.) 8 February 1990 see the whole document --	35-47	A	EP, A, 0297379 (MOLECULAR DIAGNOSTICS) 4 January 1989 --		A	Biotechniques, volume 8, January 1990, K.H. Roux et al.: "A strategy for single site PCR amplification of dsDNA: Priming digested cloned or genomic DNA from an anchor-modified restriction site and a short internal sequence", pages 48-57 cited in the application --		A	Gene, volume 84, 1989, Elsevier Science Publishers B.V. ./.	
Category <sup>9</sup>	Citation of Document, <sup>11</sup> with Indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>																		
A	EP, A, 0356021 (IMPERIAL CHEMICAL INDUST.) 28 February 1990 see the whole document cited in the application --	1-34																		
A	WO, A, 90/01065 (GENELABS INC.) 8 February 1990 see the whole document --	35-47																		
A	EP, A, 0297379 (MOLECULAR DIAGNOSTICS) 4 January 1989 --																			
A	Biotechniques, volume 8, January 1990, K.H. Roux et al.: "A strategy for single site PCR amplification of dsDNA: Priming digested cloned or genomic DNA from an anchor-modified restriction site and a short internal sequence", pages 48-57 cited in the application --																			
A	Gene, volume 84, 1989, Elsevier Science Publishers B.V. ./.																			
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p><sup>10</sup> Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"A" document member of the same patent family</p> </div> </div>																				
<b>IV. CERTIFICATE</b>																				
Date of the Actual Completion of the International Search <div style="text-align: center; margin-top: 10px;">25th July 1991</div>	Date of Mailing of this International Search Report <div style="text-align: center; margin-top: 10px;">11.09.91</div>																			
International Searching Authority <div style="text-align: center; margin-top: 10px;">EUROPEAN PATENT OFFICE</div>	Signature of Authorized Officer <div style="text-align: center; margin-top: 10px;">          miss T. MORTENSEN       </div>																			



III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, " with indication, where appropriate, of the relevant passages	Relevant to Claim No.
	V. Shyamala et al.: "Genome walking by single-specific-primer polymerase chain reaction: SSP-PCR", pages 1-8 cited in the application --	
A	Biochemical and Biophysical Research Communications, volume 167, no. 2, 16 March 1990, Academic Press Inc. M. Kalman et al.: "Polymerase chain reaction (PCR) amplification with a single specific primer", pages 504-506 cited in the application --	
A	Proc. Natl. Acad. Sci., volume 86, August 1989, Biochemistry O. Ohara et al.: "One-sided polymerase chain reaction: the amplification of cDNA", pages 5673-5677 --	
P,X	Nucleic Acids Research, volume 18, no. 10, 25 May 1990, (Eynsham, Oxford, GB) A. Rosenthal et al.: "Genomic walking and sequencing by oligo-cassette mediated polymerase chain reaction", pages 3095-3096  -----	1-34

**ANNEX TO THE INTERNATIONAL SEARCH REPORT  
ON INTERNATIONAL PATENT APPLICATION NO.**

GB 9100803  
SA 47500

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 02/09/91  
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0356021	28-02-90	AU-A- 3896689	22-03-90
		GB-A- 2221909	21-02-90
		JP-A- 2174679	06-07-90
WO-A- 9001065	08-02-90	AU-A- 4049589	19-02-90
		EP-A- 0426756	15-05-91
EP-A- 0297379	04-01-89	JP-A- 1098499	17-04-89